

Journal of Abnormal Psychology

Monograph

Vol. 76, No. 3, Part 2

December 1970

HYPNOTIC AGE REGRESSION: AN EMPIRICAL AND METHODOLOGICAL ANALYSIS¹

DONALD N. O'CONNELL,² RONALD E. SHOR,³ AND MARTIN T. ORNE⁴

Institute of the Pennsylvania Hospital and University of Pennsylvania

A repetition was made, with additional control groups, of the study of Reiff and Scheerer of hypermnesia through reinstatement of prior cognitive modes of functioning during hypnotic age regression to ages 10, 7, and 4. Only partial replication of the original findings was obtained. Tests amenable to the influence of *E* expectation and demand characteristics showed the least replication. Tests resistant to such influences replicated well. No evidence of hypermnesia was found. Comparisons with other control (or quasi-control) groups showed: (a) equal performance of cryptosimulating *Ss* in the absence of hypnosis; (b) evidence for confounding in the original study of treatment and design effects; (c) moderate effect of presence or absence of role support in a quasi-participant group; and (d) fairly good behavioral validity in direct comparisons with children. Evidence for transcendence during hypnosis of waking role-playing behavior was lacking, although striking subjective alterations were present in hypnosis. Methodological implications of these findings are stressed.

The Use of Hypnosis as an Experimental Control Technique

The provision of adequate experimental control in psychological research has always

¹ All stages of this research and the preparation of the monograph were supported by Grant MH-03369-05 and by Grant MH-11028-01 from the National Institute of Mental Health. Revisions in the final version of the manuscript were supported in part by Grant MH-04172-10 and by Grant MH-19156-01.

The authors are indebted to their colleagues Frederick J. Evans, Lawrence A. Gustafson, Charles H. Holland, Edgar P. Nace, Emily Carota Orne, David A. Paskewitz, Campbell W. Perry, and especially A. G. Hammer for their helpful comments. Special thanks are due to Ulric Neisser for his valuable comments and statistical advice. The authors are particularly grateful to Robert Reiff for his cooperation in the replication and his help in personally rating the word associations obtained in the present study. They greatly appreciate the care exercised by Mae Weglarski and Lani L. Pyles in the preparation and typing of the manuscript.

² Now at Massachusetts Mental Health Center.

³ Now at the University of New Hampshire.

⁴ Requests for reprints should be sent to Martin T. Orne, Unit for Experimental Psychiatry, Institute of

been a particularly vexing methodological problem, since the very participation of an *S* in an experiment produces changes in him that are irreversible. This problem has been met by the use of matched groups, balanced order designs, and a wide variety of similar procedures, including at times the use of identical twins. An ideal fantasy solution would be provided by a matter-duplicator allowing *E* to obtain as many exact copies of his *S* as required. In the absence of this magical apparatus, some investigators have made do with second best by enlisting the aid of hypnosis. It is with the substantive conclusions drawn from such an application of hypnosis and the methodological implications following therefrom that the present study is concerned.

This use of hypnosis has been justified by the assumption that hypnotically induced amnesia produces a true temporary ablation of the amnesic material. If this assumption is

the Pennsylvania Hospital, 111 North Forty-Ninth Street, Philadelphia, Pennsylvania 19139.

correct, hypnotic amnesia would readily provide a means of reusing Ss under several treatment conditions without contamination from one condition to another. An example of an experiment of this type is the study of the dependence of coin-size estimations by children on their socioeconomic status (Bruner & Goodman, 1947) made by Ashley, Harper, and Runyon (1951), in which undergraduate Ss were age regressed and given artificial histories. An empirical evaluation of the legitimacy of this procedure has been made by Orne (1959).

The apparent validity of hypnotic age regression has even lured some investigators into doing short-cut longitudinal studies with its aid. One example is the analysis of the development of vocational interests made by Kline and his co-workers (Kline, 1953; Kline & Haggerty, 1953; Kline & Schneck, 1950). A frequently cited (though seldom read)⁵ example is the case study of the development of personality structure from birth, reported by Gakkebusch and his colleagues (Gakkebusch, 1928; Gakkebusch, Polinkovskii, & Fundiller, 1930). The certainty with which hypnotic age regression was viewed as valid is evidenced in this study by the authors' suggestion that the technique would be of value to genetic reflexology.

It is obvious, however, that justification for the use of hypnotic techniques as research tools with which answers to basic psychological questions not intrinsically related to hypnosis can be obtained is strongly dependent upon the reality of the hypnotic phenomena being induced. Yet recent reviews of hypnotic age regression (Barber, 1962; Gebhard, 1961; Yates, 1961) are in agreement that evidence for the reality of regression phenomena, in the sense that they transcend the range of normal performance and recall, is scant indeed.

As LeCron (1948) has pointed out, many Es, as well as clinicians, have found hypnotic age regression to be one of the most interesting and seemingly central phenomena found in hypnosis. Regressions of the type wherein the person is reliving a past experience and acting it out in detail, often with great affect, are

particularly striking. This type of regression apparently requires the mobilization of a very wide range of deep trance phenomena, including positive and negative hallucinations in several sense modalities, temporary amnesia for present events, hypermnesia for past events, and even physiological changes. Age regressions of this type are one of the few phenomena that reproduce in the experimental laboratory the affective intensity found in psychotherapeutic interactions.

Certain tests have been assumed not amenable to simulation, and success on these tests has provided the strongest evidence for the validity of age regression. An example is the ability of age-regressed Ss to identify correctly the days of the week of birthdays and Christmas days in their childhood (True, 1949). Attempts to replicate this finding, however, have failed when adequate precautions were taken to prevent the possibility of Ss obtaining the required information by devious means (Barber, 1962). The assumption that the average person cannot give immediate and accurate recall to such material seems justified by everyday experience.

An important feature of the testing procedure used by True (1962) is the fact that *E*, when testing, was aware of the actual day of the week upon which *S*'s former birthday fell and asked in progression, "Was it Monday? Was it Tuesday? Was it Wednesday? . . .," and so on. This provides an excellent opportunity for *E* unconsciously to communicate appropriate verbal cues to *S*, which *S* can then utilize to improve his number of correct hits.

Certain other types of tests have been used under the presumption that they cannot be simulated because knowledge of the differences between adult and preadult performance is available only to the expert. Examples are the Rorschach and Word Association Tests. However, empirical tests of the assumption of difficulty or impossibility of the simulation of children's response patterns have usually not been made.

Orne (1959) has demonstrated the extreme difficulty of finding behavioral criteria of the presence of hypnosis that cannot be matched by suitably motivated simulating Ss. This finding raises serious doubts about previous

⁵ For the astonishing history of the citation of this reference, see the footnote on p. 143 of the review of Gebhard (1961).

attempts to demonstrate the validity of hypnotic age regression. Until suitable tests proven not to be open to simulation are used and shown to be passed with success by age-regressed Ss but not by equally motivated control groups, these doubts will remain unresolved. Until the validity of hypnotic age regression is demonstrated, its use as a research tool cannot be justified.

In order to test the validity of hypnotic age regression and demonstrate the need for suitable methodology in the evaluation of hypnotic phenomena in general, it was felt desirable to attempt the replication of an outstanding recent investigation of hypnotic age regression, and to elaborate the original design of such a study to include the types of control groups needed to clarify the interpretation of the results obtained.

The Work of Reiff and Scheerer

Perhaps the most outstanding recent example of the application of hypnotic procedures for the investigation of psychological variables has been the investigation of memory function via hypnotic age regression made by Robert Reiff (Reiff, 1954) and later published in conjunction with the late Martin Scheerer as a monograph, *Memory and hypnotic age regression: Developmental aspects of cognitive function explored through hypnosis* (Reiff & Scheerer, 1959). Because of the widespread interest shown in this monograph, the methodological care exhibited in its design, and the theoretical implications of its results, a replication of this work was chosen as eminently suitable for this purpose.

As has been previously pointed out (Orne & O'Connell, 1961), the monograph may be treated as two related but separable sections: (a) a presentation and elaboration of a theory of the development of cognitive functioning and its relation to memory and (b) a carefully planned experimental test of this theoretical position. The present study was only tangentially concerned with the first portion of the monograph, but was directly concerned with the suitability of the experimental procedures used for testing it.

The hypothesis being tested by Reiff and Scheerer was, briefly, that the availability of

specific memories for past events is a function not only of their distance in time but also of the mode of cognitive functioning used at the time they occurred. The work of Piaget and others has produced evidence that cognitive functioning undergoes marked developmental changes and that qualitatively different levels of functioning can be differentiated. As the child grows older, his mode of thinking changes from autistic and egocentric to realistic and rational modes of thought.

A distinction was drawn between two forms of memory: (a) *remembrances*, which were defined as memories having an experienced personal-temporal or autobiographical index; and (b) *memoria*, defined as memories lacking the experience of such an index. Remembrances included such memories as being tow-headed when a child of 4, or entering school in Chicago. Memoria, on the other hand, included such memories as items from the generalized informational repertoire of the individual, automatized habits, and, most relevant to the present argument, the schemata of cognitive functioning themselves. An example would be knowledge that Portland is the largest city in Oregon, or that $2 \times 2 = 4$. As the child grows into the adult, earlier schemata of cognitive functioning become increasingly unavailable to volitional reactivation. They may reappear under certain conditions, for example, narcosynthesis, severe emotional stress, during senile reminiscence, and presumably under hypnosis.

It was hypothesized that childhood memories which are not ordinarily available to adult recall could become so if the cognitive schemata operating when they were laid down could somehow be revived. It was proposed that this can be achieved by hypnotic age regression.

It is obvious that an adequate test of the dependence of memory availability on schemata of cognitive functioning can be made using hypnotic age regression only if hypnotic age regression can be relied upon as a valid procedure for the reinstatement of developmentally earlier modes of thinking. Reiff and Scheerer were well aware of this requirement and devoted a portion of their monograph to a critical review of relevant work in this area. They concluded that there is adequate experi-

mental evidence to warrant the use of hypnotic age regression for their experimental purpose. However, the experimental tasks chosen by them were in part chosen to demonstrate the validity of hypnotic age regression. To a certain extent, this was a bootstrap operation.

Two groups of tests were devised by them. The first group of tests was concerned with modes of emotional and cognitive functioning during hypnotic age regression. Based largely on tests used by Piaget and others, these could be scored in terms of developmental level of functioning on the autistic-egocentric-rational continuum. These tests would provide an internal check of the efficacy of hypnotic age regression in reinstating earlier cognitive schemata and would thus also provide a test of the validity of hypnotic age regression itself—a very ingenious procedure.

The second was concerned with contextual recall. They were chosen to demonstrate hypermnnesia for remembrances during hypnotic age regression. Questionnaires were administered that contained items of information that could be validated from existing school records.

Reiff and Scheerer found marked differences between the gross behavioral performances of their experimental and control Ss. They also found statistically reliable differences between the performance levels of the two groups on many of the specific tests concerned with cognitive and emotional functioning. These differences were in the directions predicted, and were interpreted as lending strong support to the validity of the hypnotic age-regression phenomenon itself.

Along with these differences in performance level, there were marked differences in the tests of contextual recall. The age-regressed Ss gave many more specific answers to the questionnaire items, whereas the control Ss tended to answer more frequently with "I don't know" or a similar evasion. The contrast between the two groups was so great that the veridicality of the answers given by the control groups was not checked by going to their school records, since so few specific answers had been given. On the other hand, a careful check of school records revealed a high incidence of correct recall among the experimental Ss.

Considerable anecdotal evidence was also presented of the recall of specific skills, for example, reading Latin, among this group which they claimed were not available to them in the normal waking state. This was also interpreted by the authors as evidence for the validity of the age-regression phenomenon.

Taken together, the results of the performance tests and the contextual recall tests were interpreted as supporting the validity of the hypotheses concerning memory function which were the primary interest of the investigation, namely, that earlier functional schemata had been made accessible to Ss by means of hypnotic age regression, and that as a consequence of this functional revivification, previously unavailable memories had become available to recall.

As preliminary reviews of this study have pointed out (O'Connell, 1961; Orne & O'Connell, 1961), there are implications of the experimental procedures and design used which cast very serious doubt upon the justification of such conclusions. Before discussing these in detail, it is necessary to outline briefly the design used by Reiff and Scheerer. They chose 5 good hypnotic Ss from among a sample of over 100 students who were initially tested for suggestibility. These constituted the experimental group. They had been screened for hypnotizability prior to their participation in the experiment and had been tested for their ability to respond to suggestions of age regression. They were age regressed to three ages: 10, 7, and 4. After each regression they were brought back to their true chronological age with suggestions of complete amnesia for their regressed experiences. The three age regressions were not made on the same experimental day.

Three control groups of Ss not selected for hypnotizability were used. Each group was asked to role-play one of the three ages to which the experimental Ss had been regressed.

This design introduces a number of disquieting factors which suggest alternative explanations for the results obtained. An initial factor is S motivation. When Ss are asked to serve as controls in a situation in which they know that they are acting as controls and only play-acting the role and further know that E is aware of this, they tend to give

strikingly unconvincing performances quite unlike the behavior of actually hypnotized Ss. In contrast, if *E* is unaware that *S* is not actually hypnotized or if there is an even chance as far as *E* is concerned that this is the case, and if in addition *S* is aware of this and has been told that it is possible for him to put one over, as it were, on *E*, then a very different order of behavior appears, which in well-motivated Ss can be quite indistinguishable from that of actual hypnotized Ss (Orne, 1959).

It is somewhat unclear whether the control Ss were or were not aware that they were serving as controls in an experiment involving hypnosis. There is a discrepancy between the instructions for the control Ss as reported in the thesis (Reiff, 1954) and as reported in the monograph (Reiff & Scheerer, 1959). The instructions used by Reiff and Scheerer to their simulating control Ss were reproduced in the thesis as follows (*italics have been added to the portion omitted in the book*):

Before we begin the actual hypnosis experiment, there are certain tests I would like to give to help me determine whether or not you will make a good subject. I would therefore like you to cooperate with me in these tasks to the best of your ability. I am going to take you into a playroom where there are lots of toys. As long as you are in that playroom, no matter who talks to you, or what you are asked to do, I would like you to pretend that you are (age) years old. When you walk into that playroom pretend that it is the day of your (age) birthday. Think like you did when you were (age) years old. Walk like you did when you were (age) years old. And talk like you did when you were (age) years old.

I would like you to continue this until we leave the playroom. At no time step out of your role. Try to be (age) years old every second you are there. Forget everything you have learned since you were (age) years old. If any questions are asked you, try to answer them as you would have answered them when you were (age) years old. When we go into the playroom, I would like you to begin to play with the toys like you were (age) years old and continue playing until I ask you to do something else. Any questions?

The experimenter then answered any questions. And if there were no questions, he continued,

Start putting yourself into the frame of mind of a (age) year old. Now we are going back to the time when you were (age) years old. Let's go into the playroom [p. 70].

Let us assume that in this instance the control Ss did ascertain that they were serving as

controls. It would not be difficult to explain their behavior as being depressed from the maximum level of role-playing behavior of which they might be capable. It is difficult for Ss to give convincing performances under such conditions. They appear self-conscious and tend to step out of role. One of the main reasons for this is that *S* is quite aware that the *E* knows that he is only play-acting and not really hypnotized. Furthermore, the working hypothesis of the *E* is that *S* will indeed not behave convincingly, and this expectation (conscious or unconscious as the case may be) is communicated to *S* by cues from *E*. The relationship is not a convincing *folie à deux*.

One can conclude from these considerations that differences in behavior found between actually hypnotized experimental groups and simulating control groups may be due in large part to conflicts of motivation to do a convincing acting job on the part of the control Ss coupled with a desire to underplay the role in order to fulfill the expectations of *E* that there will be large differences in performance between real and control Ss. The *S* acts as he feels *E* expects him to act.

A second factor which could have produced the results of Reiff and Scheerer is introduced by the design itself. The demand characteristics of the design are such as to produce differences in performance between the experimental and control groups even if the motivational factors discussed above were not present. Many of the tests used at age 10 were repeated again in identical form at age 7 and even again at age 4 in some cases. The experimental Ss were successively regressed to all three ages, but the control Ss were asked to act out only one age.

Two arguments were put forward in defense of this procedure: (*a*) that there would be no interaction from one age to another in the experimental group since complete amnesia was induced after each age regression and would act as a barrier between ages; and (*b*) that the extreme difficulty of obtaining really adequate hypnotic Ss made impractical the use of three different experimental groups, each age regressed to only one age.

While one can certainly sympathize with the second argument, it is not an adequate reason for using only the control groups that

were used. A group of control Ss could easily have been included which acted out all three ages in a manner analogous to that done by the experimental group.

The first reason, however, is open to a more serious criticism. There is every evidence that hypnotic amnesia is not functionally complete, that is, that it is not analogous to an ablation of previous experience, but rather only to an active repression of that experience. Previous experiences while age regressed, even though not remembered consciously, could influence performance on later age regression. This becomes apparent when one considers that just such an effect must take place when a posthypnotic suggestion is carried out in the waking state with complete posthypnotic amnesia. Memory for the suggestion, which was given during hypnosis, must be operative at some level of psychic organization in order for the posthypnotic suggestion to be effective when the cue is given by *E*. There is, in addition to this logical evidence, ample experimental evidence demonstrating that a repression model rather than an ablation model fits the facts of hypnotic amnesia, for example, the work of Strickler (1929) as reported by Hull (1933) or the elegant study of Leonard (1965).

Since many of the individual tests were exactly repeated at consecutive ages, the experimental Ss' experiences with them (whether consciously available or not) must indeed have influenced their performance when the tests were presented again. This influence could occur in at least two different ways. Most adults will have some vague idea of the levels of performance appropriate to ages 10, 7, and 4. They are aware, for example, that 4 is a preschool age, and that therefore the child will be unable to read. At least they should be able to work on the rationale that as one progresses to earlier and earlier ages, there is a progressive simplification in performance. At the earliest age, age 4, they might assume that most children would be totally or partially unable to perform at all on the tasks given them in the test battery.

In addition to such vaguely formulated criteria, they may in fact have occasion to observe appropriate behavior with younger re-

lations or the children of friends and thus have some bench marks by which to judge.

The second effect is more subtle and could be operative in instances in which no idea at all was available of what behavior would be appropriate. This would be particularly relevant on the Word Association Test, where such information is certainly not common knowledge. In this case, even though they might not know what responses were appropriate to earlier ages, they would be very likely to assume that the responses at, say, age 10 would be somehow different from those at the adult level; that the responses of a 7-yr.-old would again be different from those of a 10-yr.-old, etc. Since the behavioral repertoire available to the adult *S* must be limited, changes from adult patterns might very well be in the direction of correct performance at earlier ages. This effect would be operative for both hypnotized and nonhypnotized Ss, provided all three ages were tested.⁶

A third factor which may have influenced the results is the procedure used for the selection of Ss. In order to meet the selection criteria for inclusion in the experimental group, Ss had to be able successfully to experience hypnotic age regression. This had to be tested in preliminary screening sessions. During these sessions, there was an opportunity for two things to occur: (*a*) for *S* to receive training in age regression and (*b*) for the *E* unconsciously to influence *S* to produce the type of performance desired. In instances where the age-regression performance was not in the desired direction, *S* may have been judged inadequate and rejected. This type of *E* bias in *S* selection could, of course, occur quite unknown to the *E* himself.

⁶ An additional factor which might contribute to the differences found between experimental and control Ss is the presence of the hypnotic state. It has been suggested that an increase in autistic thinking is a concomitant of the hypnotic state itself (Meares, 1960). Since autistic thinking is also typical of early developmental stages in the intellectual growth of the individual, hypnotically age-regressed Ss might appear to be functioning in this manner not as the result of the suggestion of age regression but rather as a result of hypnosis itself. Neither the original monograph nor the present one was designed to investigate this possibility, so it will not be considered further.

A fourth factor may be intrinsic differences in the two populations sampled. Although no data are given on the hypnotizability of the control Ss, we may assume that they were in the low to medium range of the hypnotizability scale. Although many reports have appeared in the literature purporting to have found personality differences between hypnotizable and unhypnotizable individuals, these have generally been conflicting and unreliable in replication studies (Hilgard, 1965; Shor, Orne, & O'Connell, 1966). Yet the suspicion persists among workers in the field that there may be differences in personality or mode of cognitive functioning that correlate with hypnotizability. The ability to return, say, to earlier patterns of cognitive functioning, may be higher in the highly hypnotizable population than in the nonhypnotizable one.

The present experiment was designed to take into account these factors and, insofar as feasible, introduce suitable controls for them. The rationale for the procedures used is given as the experimental design is described. A summary of the present experimental design is given in Table 1.

METHOD

The Real-Simulator Design

A major concern of the present study was to determine to what extent, if any, the results obtained by Reiff and Scheerer were truly due to the re-introduction through hypnotic techniques of cognitive patterns of functioning unavailable in the absence of the hypnotic state. To put this another way, could these results have been produced under suitable motivation by Ss who were not hypnotized? The real-simulator design was developed specifically to answer this type of question (Orne, 1959).

In the real-simulator design, unhypnotizable Ss are instructed to pretend to be hypnotized to the best of their ability when participating in an experiment with an *E* to whom they are previously unknown. They are informed that this *E* is aware that some of the Ss who are participating in the experiment are not really hypnotized but are only pretending, but that he is not aware which Ss fall into this category. They are further told that if at any time *E* becomes aware of the fact that they are only pretending, he will stop the experiment. Therefore, as long as the experiment continues, they can rest assured that they are doing a satisfactory job. They are told that it is very difficult to pretend in a convincing manner that one is hypnotized, but that it is possible. The *E* is indeed aware that some of the Ss are truly hypnotized whereas others are

TABLE 1
SUMMARY OF EXPERIMENTAL DESIGN

Group	N	Tasks			
		Adult	Age 10	Age 7	Age 4
Age regressed	10	X	X	X	X
Cryptosimulators	10	X	X	X	X
Three-age role players	10	X	X	X	X
One-age role players					
Age 10	10	X	X	—	—
Age 7	10	X	—	X	—
Age 4	10	X	—	—	X
Quasi-participants	10	X	X	X	X
Children					
Age 10	10	—	X	—	—
Age 7	10	—	—	X	—
Age 4	10	—	—	—	X
Total	100				

only pretending, but he is not aware of which Ss fall into which category, nor is he told at any time. It should be stressed that *S* is not coached or given any training on the tasks to be performed while feigning hypnosis. He must rely on his own knowledge, guesses, or imagination of what behavior will be appropriate to a hypnotized *S*. Such information is also available to the hypnotized *S*. The assumption being tested is that being in the hypnotic state will result in additional behavior or modifications of behavior which supersede that due to the expectations of *S* and *E* alone.

This procedure maximizes the motivation of the pretending *S*. It is a situation in which he has the double satisfaction of successfully playing a difficult role and at the same time of, as it were, putting one over on *E*. Further, since *E* is not aware of which Ss fall into which category, unless he is able successfully to detect the pretending Ss, he will treat some Ss in each group as real hypnotic Ss to the extent to which he is convinced that they fall into this category. Prior experiments in which this design has been used have shown that without special procedures, even experienced hypnotists are unable in this situation to detect the *S* who is pretending to be hypnotized with any better than chance accuracy. A comparison of the hypnotized and the pretending groups allows one to see to what extent, if any, the behavioral repertoire of *S* has been enlarged by the introduction of hypnosis.

Experimental and Control Groups

Introduction of the real-simulator design complicates somewhat the attempt to replicate Reiff and Scheerer's results. The attitude of *E* in this type of experiment is not identical to that of *E* who has no reason to doubt that his hypnotic Ss are truly hypnotized. In order to be strictly comparable with the Reiff and Scheerer experiment, it might have been desirable to run 10 Ss who were truly deeply hypnotizable Ss and were known as such to *E*. Again,

however, economic considerations precluded the inclusion of such a group.

The imbalance in the original study, arising from the fact that experimental (hypnotic) Ss were consecutively tested at all three stages, whereas control groups were tested at only one age each, could be eliminated in either of two ways, either by use of a control group tested at all three ages or by use of three hypnotized groups tested at one age each. This latter possibility, however, is a most unparsimonious use of good hypnotic Ss, for in order to obtain one extremely hypnotizable S, very large numbers of potential Ss must be screened.

In the present study, it was decided to use 10 rather than 5 hypnotic Ss in the group corresponding to Reiff and Scheerer's experimental group. For reasons of economy, it was decided to adopt the alternative of including a control group required to role-play all three ages as well as three control groups corresponding to those used in the original experiment. A comparison of these two types of control groups provides information on the magnitude of the effect of repetitive testing.

Two types of factors are operative in the usual hypnotic experiment, treatment variables and *E*-bias variables. Frequently the two are confounded. In a typical hypnotic experiment, an experimental group of Ss known to be hypnotizable is used and their behavior compared to that of a control group which may be un hypnotizable or whose hypnotizability may be unknown at the time of the experiment. The *E* is perfectly aware at all times within which group each S falls. There is, therefore, the probability that he will react differentially to these two groups and thereby influence differentially their behavior. One advantage of the real-simulator design is that treatment effects and *E*-bias effects are in a certain sense orthogonal to one another, and their differential effects can be separately identified.

The *E*-bias variables can be directly controlled to a large extent. In the present experiment, the primary *E* (DNO'C) consciously adopted a uniform attitude to not only the experimental groups (real and pretending) but also to the one-age control groups and the three-age control group. His attitude was one of strong role support and the expectation of successful performance throughout all tests. An objective indicant of the success of this attitude of uniform role support is available in the frequency of Ss breaking out of their role. In the original study such instances were reported as frequent. In the present study no instances of this were observed. This attitude was adopted as an attempt to hold constant as much as possible the *E*-bias variables. In order further to explore the effect of these variables, and in particular of role support, an additional control group was introduced into the experiment. With this group, *E* avoided any role support so far as possible. The Ss were instructed to treat the entire experiment as an intellectual exercise, as an experiment in which they would not actually take part. They were asked to report what they thought they would do if they had indeed taken part in the ex-

periment. This type of group has been described elsewhere as the "pre-inquiry" or "non-experiment" quasi-control (Orne, 1969).

A note on terminology. The term "simulator" has been used in a variety of ways in hypnotic research. In experiments employing the real-simulator design, it has been used with the specific meaning of an un hypnotizable S who has been instructed by one *E* to pretend to be hypnotized when participating in an experiment with a second *E* who does not know whether S is indeed hypnotized or not. A second and more common usage is that adopted by Reiff and Scheerer, who used the term simulator to indicate control Ss who are asked by *E* to play the role of a child then and there. Both *E* and S are aware of the task and its implications. Notice in this usage that the S is not asked to play the role of a hypnotized S, but is asked to play a role which is comparable to that induced through hypnosis in the experimental S.

A third usage of the term simulator has been to indicate control S who is asked to pretend to be hypnotized, again then and there, with full knowledge that the *E* with whom he is pretending knows that he is only pretending to be hypnotized.

These various usages of this term have led to a great deal of confusion. The term simulator is therefore not used in the present study; rather we refer to role players to indicate control Ss who are asked to play the role of children by an *E* who is obviously aware that S is doing so. The S who is role-playing is not asked to play the role of a hypnotized S who is age regressed, but only of a child. No mention is made of hypnosis when these Ss are instructed. This corresponds to the second usage indicated above.

A special term is introduced to indicate the first usage. Those Ss who are pretending to be hypnotically age regressed in the real-simulator design are termed "cryptosimulators." This term will be used throughout to refer only to control Ss who are un hypnotizable and who are asked to role-play the part of age-regressed hypnotized Ss in the hope of fooling an *E* who is aware that some Ss will be role-playing, but not of which ones will be doing this.

Addendum. In addition to these experimental and control groups, it was felt highly desirable to include real children of ages 10, 7, and 4 as validation groups against which the other groups could be compared. This has rarely been done in research in hypnotic age regression.

Overall review of groups. It will be observed that the various groups to be studied line up along a continuum: quasi-participants, one-age role players, three-age role players, cryptosimulators, hypnotically age-regressed Ss, and children of appropriate age groups. Further, the design arranges that these groups undertake a number of activities differentially amenable to simulation. In these circumstances it is anticipated that where simulation in some sense is possible, both cryptosimulators and regressed Ss will alike approach the validating groups, while with

other procedures they will fail to do so. It is expected also that these groups will behave differently from the other simulators, whom we have referred to here as role players, suggesting how Reiff and Scheerer obtained their results. Nevertheless, it is also expected that the cryptosimulators and age-regressed Ss will "genuinely" differ in their subjective experiences. In this study every opportunity has been provided for regression to occur as fully and realistically as possible. If the results come out as anticipated, then by virtue of the combination of groups and procedures employed, they will point rather strongly to the *positive* conclusion that hypnotic age regression, although it does not involve intentional deception, is nevertheless something less than the total reinstatement of a former child-like mental state. At least the more parsimonious view would be that age-regressed behavior is more analogous to a temporary delusion (as of being Napoleon) than to "regressed" behavior.

General Criteria of Subject Selection

All adult Ss were obtained from the college undergraduate population of institutions in the Boston area, primarily Harvard University, Massachusetts Institute of Technology, Brandeis University, and Boston University. They were all volunteers, and all were paid for their participation. Important additional qualifications were: (a) that they had attended grade school entirely within the United States educational system (including Hawaii and Alaska) and (b) that English was their native language.

Regional differences between the sample in the present experiment and that used by Reiff and Scheerer are difficult to assess but may be a possibility. Distributions of hypnotizability have been found to vary significantly among samples obtained from different university populations even when tested with standardized tests of hypnotic susceptibility (Hilgard, 1965).

Treatments of Specific Experimental and Control Groups

Age regressed. Paid volunteer college undergraduates were screened for hypnotizability. Well over 400 students were screened before the required 10 deeply hypnotizable Ss were obtained. Criteria for inclusion in this group were that Ss show the usual phenomena of somnambulism, for example, complete posthypnotic amnesia, positive and negative hallucinations, posthypnotic suggestions, etc., and that they be able to experience subjectively convincing age regression. Initial screenings were made by various staff members, but all Ss were given a minimum of one (and usually several) intensive clinical diagnostic screenings by the second E (RES). They were not included in this group until he was convinced that they completely fulfilled the criteria. Since really convincing hypnotic age regression is a rare phenomenon, Ss in this sense were conservatively chosen and, we believe, they equal or exceed in hypnotizability those used in any previous experiments.

Postexperimental inquiries were made to ascertain the adequacy of the age-regression experiences for these Ss. Any Ss who failed to meet the criteria of complete subjectively convincing age regression were to be eliminated from the study. In actual fact, there was not such an instance.

In all cases the entire test battery for all three ages was given in one experimental session, which lasted from 3 to 4 hr. (The three ages were induced in decreasing chronological order, as was done in the original study.) This departs somewhat from the procedure used by Reiff and Scheerer (1959), who stated:

Sometimes a subject was tested at one and sometimes at two of the specified ages in the same hypnotic session, but he was always brought back to his chronological age after each of the regressed ages [p. 109].

Similarly, Ss in the present experiment were returned to their chronological age between test ages and intervening amnesia was induced. The Ss given all three ages of tests were run during one session in order to treat all Ss consistently and, more important, to avoid giving Ss the opportunity for covert rehearsal between experimental sessions, either deliberately or through fantasy elaboration.

Cryptosimulators. These Ss were obtained by the same screening procedure described above; however, the criterion for their inclusion in this group was lack of elicitation of hypnotic phenomena, particularly those phenomena involving subjective alterations in response to suggestion, for example, feelings of lightness in response to suggestions of arm levitation. Final acceptance screenings and postexperimental inquiries for these Ss were also made by the second E.

In terms of the diagnostic rating system described by Orne and O'Connell (1967), the age-regressed Ss rated 5 or 5+, and the cryptosimulating Ss received ratings no higher than 2--.

In accord with the real-simulator design chosen, the age-regressed Ss and the cryptosimulating Ss were presented to E in a randomized order. It was agreed that Ss would be eliminated from the study if the primary E were convinced that they were cryptosimulating. (This happened in one instance.)

Three-age role players. These Ss were solicited from the undergraduate population by advertisements asking them to volunteer to take part in "a psychological experiment," the nature of which was unspecified. Care was taken that they not be given information that would lead them to expect that hypnosis would be in any way involved in the experiment for which they had volunteered. They were given instructions to play the role of children of the ages described. They were not asked to pretend that they were hypnotically age regressed. The instructions used were those reported in the *monograph* (Reiff & Scheerer, 1959).⁷

⁷ The thesis was unavailable at the time the present experiment was designed. We were at that time

They were asked to play the role of first a 10-yr.-old, then a 7-yr.-old, and finally a 4-yr.-old. Between ages, they were taken out of role and treated as adult Ss. This was analogous to the procedure followed with age-regressed Ss in the original study. After the completion of these three tasks, they were asked by *E* what they felt the experiment was about. In no instance did they report that they had associated the experiment with hypnosis; if any *S* had done so, he would have been eliminated from the experiment. After further questioning about their feelings and experiences while playing the role of children, the true nature of the experiment was briefly explained to them and their further cooperation was requested in taking a standardized test of hypnotizability. If they agreed to do so, they were given Form A of the Stanford Hypnotic Susceptibility Scale (SHSS:A) of Weitzenhoffer and Hilgard (1959).⁸ If they were unwilling to take the Stanford scale, they were dropped from the analysis.⁹ These Ss role-played all three ages in one session.

One-age role players. This control group corresponds to the "simulator" group used in the original study of Reiff and Scheerer. It contains three subgroups, one each for ages 10, 7, and 4. Each subgroup consisting of 10 Ss was obtained as were the three-age role players from volunteers among the college population. They were instructed, however, only to play the role of a child of one age. The ages were randomly assigned. Then their hypnotizability was tested with Form A of the Stanford scale after the completion of the test battery and the postexperimental inquiry, as above.

Reiff and Scheerer (1959) reported that the Word Association Test was given to their 15 "simulator" Ss as follows: "One half was given the Word Association Test prior to simulating any age level and the other half afterwards [p. 110]." A similar procedure was followed in the present instance, 5 Ss in each age subgroup receiving the Word Association Test as adults prior to role playing and 5 receiving it after role playing, the sequence being randomly assigned within each subgroup.

Quasi-participants. In order to determine the relative extent to which the changes found among Ss' performance levels on the tests at the three age levels are a function of the close interaction between *E* and *S* and the consequent role support given *S* by the assumption of an appropriate role by *E*, an additional group of control Ss was devised from whom test performances were obtained with mini-

mization of these factors. The interest here was in the extent to which appropriate changes in behavior could be obtained from Ss in the normal waking state in the absence of the folie à deux aspects of the hypnotic situation.

Ten Ss solicited from the undergraduate population as in the role-playing control groups were selected randomly. As before, hypnosis was not mentioned until after the testing period.

They were given the following instructions at the beginning of the experiment:

We are interested in how you would respond in an experiment which I shall describe to you. You will not actually participate in the experiment. First, I would like you to take these tests.

At this point the presentation of the waking tests was made, Questionnaire A and the Word Association Test. At the completion of these tests, the preliminary instructions were resumed as follows:

Now, if you *were* a subject we would now induce hypnosis. You would have had several training sessions previously and you would have developed the ability to enter a deep hypnotic trance. In these training sessions you would have also developed the ability to have complete hypnotic amnesia—the ability to forget the events that occur while you are in hypnosis.

I would like you to regard this as an *intellectual exercise*. I do *not* want you to pretend in any way that you actually are in the experiment.

This is not a personality test.

At some points I will have to score the first answer you give in order to make the scoring comparable to that used in the real experiment. I will tell you when these tests occur, and you just give me the first answer you think of. However, if at any time you have second thoughts or comments you would like to make, please feel free to do so.

Are there any questions?

Questions were then answered by paraphrasing the instructions. The *E* then concluded the preliminary instructions thus: "All right, now I will read you the instructions you would receive if you were actually to take part in the experiment."

The trance induction procedures were then read. The *E* was very careful at all times to control the inflection of his voice so as not to give role support to *S*. Furthermore, after each question of a test, *E* made some comment along the line of "Now, what do you think you might say then?" or "What do you think you would do then if you were really in the experiment?"

At the conclusion of this mock-experimental session, *S* was asked if he would participate further by taking a test of hypnotizability. If he consented, the Stanford scale was then administered.

Children. Many of the tests used by Reiff and Scheerer, as noted previously, were derived from the work of Piaget and his co-workers. While these tests

unaware that there was a discrepancy between the instructions as reported in the monograph and those given in the thesis.

⁸ It is unfortunate that the more advanced form of this scale, Form C (Weitzenhoffer & Hilgard, 1962), was not available at the time since it has been found that an initial testing with Form A does not provide a maximal estimate of plateau hypnotizability (O'Connell, Orne, & Shor, 1966).

⁹ Only one *S* did not agree to participate and had to be dropped from further analysis.

were well chosen and suited to the problem at hand, for many of them norms were not available at the time of the experiment on samples drawn from the American elementary school population, the population to which both the experimental and control groups had belonged when they were children. In addition, on the Word Association Test no norms at all had, up to that time, been published for children of preschool age, yet this test was used at age 4. Behavioral criteria for this age level were assumed in the original study by extrapolation from ages for which norms were available. Considering the very rapid qualitative changes that occur during these earliest years of child development, this seemed a rather dubious procedure.

In order, therefore, to obtain norms, scant though they might be, on a population comparable to the population from which the adult groups were drawn, as well as to get norms using the exact testing procedures employed in the study, it was decided to include three groups of actual children, ages 10, 7, and 4. These were considered validating groups with which the behavior of other groups could be compared.

It was desirable to have a group of children which would be comparable to our adult population when they were children, namely, a group of children likely to continue their education through college levels. Children of ages 10 and 7 were obtained through the cooperation of the Lesley Ellis School, Cambridge, Massachusetts. These were children of superior intelligence (IQs ranging from 122 to 165) and therefore likely to continue on to college. The 4-yr.-old children were obtained through the cooperation of the Harvard Pre-School, the children being primarily from the homes of faculty members in the area.

In order to establish rapport with 4-yr.-olds, the primary *E* attended the Harvard Pre-School daily for a period of 2 wk. until the children had become friendly with him and had accepted him. The test battery was presented to these children as a series of games *E* would like to play with them. Children at this school were frequently tested by members of the educational and psychological departments of the University and were unusually at ease in such a situation.

The children at the Lesley Ellis School were asked to take a series of tests; their cooperation was easily obtained since they were also a rather test-wise group.

The administration of the Word Association Test to the 4-yr.-old children presented serious procedural problems requiring considerable improvisation on the part of *E*. These will be dealt with in detail when the results of the tests are discussed.

Testing Procedures

The testing procedures used in the present study were identical with those used by Reiff and Scheerer, although in some instances analyses were made of them which had not been made in the original study. They may be grouped into two categories.

The first group is made up of tests derived mainly from the extensive investigations of Piaget and others of the genetic development of cognitive processes in children. These were designed to provide information on the modes of cognitive functioning of the *Ss* during hypnotic age regression. To the extent that these modes of functioning are congruent with the cognitive processes typical of children of the age groups suggested, these tests reflect the degree of validity of the hypnotic age-regression phenomenon itself. These procedures were:

1. A Free Play Period.
2. The Hollow Tube Test (involving elementary induction).
3. Left and Right Test.
4. The Word Association Test.
5. An Arithmetic Test.
6. Writing the Pledge of Allegiance.
7. The Clock Test.
8. Eating Lollipop despite Muddy Hands.

The second group of tests is made up of a number of questionnaires designed to provide information on the amount of contextual recall obtainable in the waking state, during hypnosis, and during hypnotic age regression. The questionnaires provide an objective test of the basic hypothesis of the monograph, namely, that reinstitution of earlier modes of cognitive functioning (in this instance by means of hypnotic age regression) can and does facilitate the availability of otherwise unrecallable information. For example, one question is, "Who sits behind you in school now?" The questionnaire was arranged in three forms, covering the same ground, with Form A couched in a form suitable at the adult level, Form B at the 10-yr. level, and Form C at the 7-yr. level. Forms B and C were given orally.

Various scores were derived from these procedures. Table 3, which gives results, provides a list of groups and of variables, and underneath the variables states succinctly the methods of scoring. It will be observed that there are blanks in the table, indicating that not all techniques were used with all groups. Where groups performed at more than one level, the order was always adult, 10 yr., 7 yr., and 4 yr. Where the Free Play Period was used, it always occurred first, and next was the questionnaire in appropriate form. Some groups repeated the adult questionnaire immediately after trance induction but before regression. The only procedures carried through as adults were the questionnaire and the Word Association Test. The remaining tests followed the questionnaire in the order set out in the above list. Inquiries, and where needed, testing of hypnotic susceptibility, followed the regressed experiences.

Statistical Procedures

Intergroup analyses. The majority of the performance scores used in the present study are at an ordinal level of measurement.

The nature of these measures entails the use of nonparametric statistical analyses. In a few instances,

TABLE 2

COMPARISON OF SIGNIFICANCE LEVELS OBTAINED ON PERFORMANCE TESTS OF REIFF
AND SCHEERER'S EXPERIMENTAL AND CONTROL GROUPS WITH CORRESPONDING
ANALYSES USED IN THE PRESENT STUDY

Test	Age	Statistic (Reiff & Scheerer)	<i>p</i>	Statistic (O'Connell et al.)	<i>p</i>
Hollow Tube Test					
Performance level	10	τ_b	.05	<i>U</i>	.05
	7		.01		.005
	4		<i>ns</i>		<i>ns</i>
Left and Right Test					
Performance level	10	τ_b	<i>ns</i>	<i>U</i>	<i>ns</i>
	4		<i>ns</i>		<i>ns</i>
Word Association Test					
Abstract responses	Adult	analysis of variance		analysis of variance	
	10		<i>ns</i>		<i>ns</i>
	7		.05		.05
	4		.03		.05
			.05		.05
Arithmetic Test					
Performance level	10	τ_b	<i>ns</i>	<i>U</i>	<i>ns</i>
	7		.01		.005
Time	10	τ_b	.02		.01
	7		.02		.01
Pledge of Allegiance					
Can write	10	Fisher's exact probability	<i>ns</i>	Fisher's exact probability	<i>ns</i>
	7		.01		.01
Clock Test					
Performance level	7	τ_b	.01	<i>U</i>	.005
Mud and Lollipop Test					
Accept lollipop	4	Fisher's exact probability	.04	Fisher's exact probability	.04

Note.—All probabilities in between-group comparisons are one-tailed.

for example, the percentage of abstract responses on the Word Association Test, parametric analyses have been justified by the underlying metric and have been used.

The frequency data obtained from the questionnaires can be treated parametrically, as can the reaction times from the Word Association Test, after suitable normalizing transformation. In most instances, however, for the sake of comparability these data have also been analyzed as ordinal measures.

The nonparametric statistic chosen for the majority of the intergroup analyses is the Mann-Whitney *U* test (Siegel, 1956). As Boneau (1962) has pointed out on the basis of empirically derived criteria, this test is in many applications comparable in power-efficiency to its parametric analog, the *t* test.

As with many nonparametric statistics, there is no ready way of obtaining exact probability levels from the *U* when between-group ties are present. The problem has been well outlined by Bradley (1960). Departures from tabled probability values for *U* are, however, conservative in direction and minor in magnitude for samples of moderate or greater size even though the number of tied scores is considerable (Siegel, 1956).

The data of the original study have been reappraised with the statistical procedures used in the present analysis. Results of this reappraisal are presented in Table 2, along with results of the analysis

made by Reiff and Scheerer using rank-biserial correlation with the tau_b statistic devised by Kendall (1955).¹⁰ The same tests are found to be statistically significant by both analyses.

Obtained values of *U* were evaluated by means of the extended tables of Auble (1953). In all instances not noted otherwise, the probabilities are one-tailed. The present study is a replication of a prior one, and on that basis, as well as from theoretical considerations, the direction of expected between-group differences can be derived.

Overall performance tests. The large number of measures being analyzed in even one comparison between groups makes it highly advantageous to be able to obtain a single intergroup comparison that is based on the overall performance level of the *Ss* in each group. A bird's-eye view of results then becomes possible. In order to attain this end, a method of applying the Mann-Whitney *U* test to average individual performance levels based on ranks was devised.

In this procedure, which will be referred to as the ranked sums of ranks test, the scores of the *Ss* making up the two groups to be compared are ranked across groups for each test. This provides a

¹⁰ Actually, Reiff and Scheerer referred to the earlier edition of this work (Kendall, 1948). The two treatments are equivalent from the standpoint of the present discussion.

column of ranks for each measure. The *S* order is kept constant, and the scores are ranked consistently from test to test, so that a higher rank, for instance, represents a higher age level. The score ranks for each *S* are then summed across tests. The row sums thus obtained (the sums of ranks) are then themselves ranked across groups, and a standard *U* test is then done on these ranked sums of ranks. This provides an index of average performance level (mean rank) for each group and a statistical evaluation of the difference between groups of this index.

The procedure allows tests given at different ages to be averaged. While tests could be differentially weighted, there is little rationale for such weightings in the present test battery, so all measures have been considered as adding equally to the final index. The measures used include scores on all tests at ages 10, 7, and 4 that were used by Reiff and Scheerer and mean log reaction times on the Word Association Test (which were not reported in the original study). A detailed listing of the test scores used and an example of the ranked sums of ranks test procedure are given in Appendix C.¹¹

Experimental Predictions

From the very design of the experiment it will be apparent that a number of expectations as to its outcome were held by the *Es*, specifically as follows:

Preliminary: It was expected that the essential effect of the Reiff and Scheerer study would be reproduced and that the one-age role-playing *Ss* would do less well than either the regressed *Ss* or the cryptosimulators.

1. It was predicted that the behavior of the regressed *Ss* would be equaled by and indistinguishable from that of the cryptosimulating *Ss*. Obviously, the subjective experiences of the two groups were expected to differ strongly. This prediction implies that the primary *E* would not be able to detect the cryptosimulating *Ss* at better than chance level.

2. It was predicted that the three-age role-playing *Ss* would do better, at ages 7 and 4, than the one-age group. It was conceived as possible that their performance might be statistically indistinguishable from that of the regressed group. However, no firm prediction could be made about this. No difference was expected at age 10, since both groups were identically treated through this point.

3. It was anticipated that because of the lack of role support, the quasi-participants would do *least well* of all the adult groups.

4. It was considered highly likely that particularly at age 4, the children would produce behavior pat-

terns qualitatively different from those found in any of the adult groups.

RESULTS

Descriptive Summary

The predictions previously set out make it clear that our main interest is the magnitudes of the differences between various paired groups. It is legitimate to consider the pairs one at a time, since planned comparisons are being carried out, and predictions have been made in advance about each pair separately. In order to avoid unnecessary repetition, and to provide an overall view of the results, a summary table is first presented which gives an appropriate descriptive statistic to represent the performance of each group in each state on each variable used in the study, followed by a discussion of the relevant particular comparisons.

Table 3 thus gives the results for the age-regressed group (as adults, at age 10, at age 7, at age 4), for the cryptosimulators at the same levels, for the three-age role players at the same levels, for the one-age role players at appropriate levels, for the quasi-participants at the four levels, and for the three groups of children of ages 10, 7 and 4.

These 25 groups of results appear in the 25 rows of the table. The first five columns give the number of cases (in each instance out of a possible 10) who showed the nominated behavior in the Free Play Period. The next two columns give mean scores of the groups on the Hollow Tube Test and the Left and Right Test. In each case the scores averaged were "levels" of performance as set out in the original Reiff and Scheerer monograph. The next column gives the means of the percentages of responses on the Word Association Test that were judged "abstract," the next the means of the log reaction times to non-critical words, and the next the means of log reaction times to critical words. The next two columns deal with the Arithmetic Test, giving, respectively, means of "levels" of performance and means of time (in seconds) spent on the problems. The next four deal with the Pledge of Allegiance. The first gives the number of cases (out of 10) "able" to write, the second the percentage of cases who included "under God" (the total was not al-

¹¹ Appendixes have been deposited with the National Auxiliary Publications Service. Order Document No. 01233 from National Auxiliary Publications Service of the American Society for Information Science, c/o Information Sciences, Inc., 909 Third Avenue, New York, New York 10022. Remit in advance \$12.00 for photocopies or \$4.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc..

TABLE 3
PERFORMANCES OF ALL GROUPS AT VARIOUS

Group	Free Play— enters readily (no. of cases)	Free Play— talks to E (no. of cases)	Free Play— talks to self (no. of cases)	Free Play— silent (no. of cases)	Free Play— in sandbox (no. of cases)	Hollow Tube (M level)	Left and Right (M level)	Word Associa- tion— Abstract responses (M %)	Word Associa- tion— noncritical words (M of log RT)
Age regressed Adults	—	—	—	—	—	—	—	59	1.07
Age 10	7	7	1	3	2	5.7	5.0	41	1.23
Age 7	9	6	2	4	5	4.0	—	26	1.38
Age 4	10	5	2	5	5	3.1	3.3	13	1.53
Cryptosimulators Adults	—	—	—	—	—	—	—	49	1.13
Age 10	6	5	2	3	3	5.0	4.9	38	1.28
Age 7	9	6	1	3	8	3.9	—	25	1.42
Age 4	9	5	0	5	8	2.2	3.3	20	1.53
Three-age role players Adults	—	—	—	—	—	—	—	44	1.09
Age 10	8	1	4	5	9	5.9	5.0	33	1.29
Age 7	9	3	3	5	7	3.9	—	19	1.31
Age 4	9	4	3	4	10	3.1	3.5	18	1.47
10-yr. role players Adults	—	—	—	—	—	—	—	51	1.05
Age 10	4	3	3	6	5	5.2	5.0	49	1.13
7-yr. role players Adults	—	—	—	—	—	—	—	34	1.24
Age 7	6	4	5	3	9	5.2	—	39	1.24
4-yr. role players Adults	—	—	—	—	—	—	—	53	1.07
Age 4	4	0	0	10	7	4.1	4.2	21	1.33
Quasi-participants Adults	—	—	—	—	—	—	—	49	1.15
Age 10	—	6	0	4	2	5.7	5.0	46	1.15
Age 7	—	9	1	0	5	4.8	—	37	1.17
Age 4	—	4	1	5	6	3.7	3.5	25	1.28
Children Age 10	7	3	1	6	4	5.4	5.0	29	1.18
Age 7	10	4	1	5	3	4.6	—	34	1.41
Age 4	10	9	1	1	7	1.4	3.5	9	

Note.—For the "free play—enters readily" procedure, data for quasi-participants were not applicable. On the Word Association Test, for both critical and noncritical words data were not taken for children aged 4.

* Double entries are used in this column for data given to Ss in both waking (above) and hypnotic (below) states.

AGE LEVELS ON ALL PROCEDURES

Word Association— Critical words (<i>M</i> of log RT)	Arithmetic level (<i>M</i> level)	Arithmetic time (<i>M</i> time)	Pledge— able to write (no. of cases)	Pledge— uses "under God" (% cases)	Pledge— words attempted (<i>M</i> no. of words)	Pledge— mis-spellings (<i>M</i> %)	Clock Test (<i>M</i> level)	Mud and Lollipop Test (no. of cases)	Adult questionnaire— A form (proportion of credible responses)	10-yr. questionnaire— B form (proportion of credible responses)	7-yr. questionnaire— C form (proportion of credible responses)
1.21	—	—	—	—	—	—	—	—	.528 ^a .662	—	—
1.56	5.3	17.8	9	0	24.3	15	—	—	—	.807	—
1.65	3.1	39.8	9	0	15.2	61	5.3	—	—	—	—
1.76	—	—	—	—	—	—	2.1	10	—	—	.786
1.22	—	—	—	—	—	—	—	—	.530 .759	—	—
1.52	4.9	20.2	10	10	25.7	14	—	—	—	.804	—
1.55	3.5	39.0	10	17	20.5	27	5.5	—	—	—	.849
1.68	—	—	—	—	—	—	2.9	10	—	—	—
1.17	—	—	—	—	—	—	—	—	.546	—	—
1.46	5.5	14.1	10	20	26.6	5	—	—	—	.588	—
1.39	4.0	23.6	8	29	18.8	32	5.5	—	—	—	.711
1.51	—	—	—	—	—	—	3.3	10	—	—	—
1.13	—	—	—	—	—	—	—	—	.514	—	—
1.33	5.0	10.0	10	50	22.5	8	—	—	—	.700	—
1.24	—	—	—	—	—	—	—	—	.634	—	—
1.33	5.4	14.1	10	29	21.6	15	6.0	—	—	—	.726
1.15	—	—	—	—	—	—	—	—	.400	—	—
1.49	—	—	—	—	—	—	3.7	8	—	—	—
1.26	—	—	—	—	—	—	—	—	.566 .636	—	—
1.19	5.6	8.7	9	0	22.7	11	—	—	—	.712	—
1.25	4.7	15.6	6	0	15.5	9	5.5	—	—	—	.726
1.19	—	—	—	—	—	—	2.4	10	—	—	—
1.34	5.3	7.9	10	88	25.8	35	—	—	—	.822	—
1.56	4.7	15.9	8	67	15.2	57	4.5	—	—	—	.986
	—	—	—	—	—	—	1.2	9	—	—	—

ways 10 since some records were unscorable for this item), the third the means of numbers of words attempted, and the fourth the means of the percentages of misspelled words. Then the next column gives the means on the Clock Test, scored in terms of levels, and the next, concluding the "performance tests," the number of cases out of 10 accepting the lollipop in the Mud and Lollipop Test.

The final three columns deal with the questionnaire responses: the first with responses as adults to the adult form, which is labeled A, the next with "10-yr. responses" to the 10-yr. form, labeled B, and the last with "7-yr. responses" to the 7-yr. form, labeled C. All entries in these columns are the proportions of responses to the relevant items (over all Ss in a group) that were credible, as defined in the text. It should be noted that in three cells in the third from the last column there are double entries, since the adult questionnaire was here given to Ss in both the waking and the hypnotic state. It should also be noted that there are numerous vacant cells since not all procedures were applied to all groups.

Performance Tests

Replication of the original study. The groups in the present study analogous to those used by Reiff and Scheerer are the age regressed ("Hypno") and one-age role players ("1-Age"), corresponding, respectively, to their experimental ("Exper") and control ("Contr") groups. Intercomparison of these groups provides a gauge of the extent to which the present study replicates the original. Certain aspects of the present design, as detailed in the preceding section, were known to differ from the original study. The expectation, nevertheless, was that the differences in performance found by Reiff and Scheerer would be sufficiently robust to appear under the conditions of the present experiment.

A number of differences between the two experiments that could affect replication can be listed:

1. The primary E (DNO'C) in the present experiment was watchfully aware of the possibly important influence of E expectations on the behavior obtained from Ss (Orne, 1959; Rosenthal, 1966). He consciously strove to treat all groups (except the quasi-partici-

pants) with equal encouragement and high role support. The extent to which this condition differed from the original experiment is not known.

2. The real-simulator design used produced uncertainty in E of the true group membership of any S presented to him for hypnotic induction, since half of these ostensibly hypnotizable Ss were in fact unhypnotizable cryptosimulators. Reiff, on the other hand, had no reason to suspect that any of his experimental group might be pretending to be hypnotized.

3. In the present experiment, all Ss were given the entire battery of tasks they were to receive in a single experimental session. In contrast, the experimental Ss in the original study were not given all three ages on a single day. The effects of intervening days may have been important, particularly for their apparent recall of questionnaire material. Barber (1962) has carefully summarized the importance of this factor on apparent accuracy of hypnotic recall, particularly in reference to apparent recall of days of the week of previous birthdays and holidays.

Detailed intergroup comparisons, test by test, have been omitted for the sake of brevity and ease of reading from the published text of the present experiment. They can be found in Appendix A, which is obtainable through the National Auxiliary Publications Service (see Footnote 11). Only the general results and specific tests showing significant intergroup differences are described here.

The score distributions upon which intergroup test comparisons are based are available in Appendix B, also obtainable through the National Auxiliary Publications Service (see Footnote 11).

Comparisons of the significance levels obtained from these intergroup comparisons are summarized, along with the corresponding significance levels obtained by Reiff and Scheerer, in Table 4, thus providing a bird's-eye view of the comparable overall results of both experiments.

Comparisons of Word Association Test measures obtained from waking presentation show no significant differences between the age-regressed group and the one-age role players who role-played ages 10 and 4. This

TABLE 4

COMPARISON OF SIGNIFICANCE LEVELS OBTAINED WITH PERFORMANCE TEST MEASURES
BETWEEN AGE-REGRESSED (EXPERIMENTAL) AND ONE-AGE ROLE-PLAYERS (CONTROL)
GROUPS IN THE PRESENT STUDY AND THAT OF REIFF AND SCHEERER

Test	Age	Reiff & Scheerer	O'Connell et al.	Statistic
		<i>p</i>	<i>p</i>	
Hollow Tube Test				
Performance level	10	.05	<i>ns</i>	<i>U</i>
	7	.005	.01	<i>U</i>
	4	<i>ns</i>	<i>ns</i>	<i>U</i>
Left and Right Test				
Performance level	10	<i>ns</i>	<i>ns</i>	<i>U</i>
	4	<i>ns</i>	<i>ns</i>	<i>U</i>
Word Association Test				
Abstract responses	Adult	<i>ns</i>	<i>ns</i>	Kruskal-Wallis analysis of variance
	10	.05	<i>ns</i>	<i>U</i>
	7	.05	<i>ns</i>	<i>U</i>
	4	.05	<i>ns</i>	<i>U</i>
Arithmetic Test				
Performance level	10	<i>ns</i>	<i>ns</i>	<i>U</i>
	7	.005	.001	<i>U</i>
Time	10	.01	.05	<i>U</i>
	7	.01	.001	<i>U</i>
Pledge of Allegiance				
Can write	7	<i>ns</i>	<i>ns</i>	Fisher's exact probability
	4	.01	<i>ns</i>	Fisher's exact probability
Clock Test				
Performance level	7	.005	<i>ns</i>	<i>U</i>
	4	<i>ns</i>	<i>ns</i>	<i>U</i>
Mud and Lollipop Test				
Accept lollipop	4	.04	<i>ns</i>	Fisher's exact probability

Note: All *p* values are one-tailed except those for the Adult comparisons, which are two-tailed.

is consistent with the findings of Reiff and Scheerer. Significant differences were, however, found between the age-regressed group and those Ss who role-played age 7. There is no a priori reason to expect such differences, and we are inclined to regard them as sampling errors, particularly taken in conjunction with other comparisons at the adult level, none of which shows significant differences. A Kruskal-Wallis analysis of variance by ranks (Siegel, 1956) showed no significant overall differences among the groups that were tested at the adult level ($H_c = 10.6$, *ns*).

Free Play Period. Reiff and Scheerer reported marked behavioral differences between their experimental and control groups during the Free Play Period at all ages, based on clinical evaluation. They reported that the control Ss seemed less at ease, stepped out of role more often, and interacted with *E* less often than did the experimental Ss. They also reported restriction of movement and lack of

appropriate voice changes in their control Ss. These observations are difficult to quantify exactly. In the present study, five relatively objective criteria of behavior in this situation were used; they were: play patterns, choice of toys, readiness to enter into the play situation, and, particularly, presence or absence of spontaneous verbal interaction with *E*. On none of these criteria were significant differences found between the age-regressed and role-playing Ss at any of the three ages. The overall clinical impression of *E* confirmed these findings, although it must be remembered that for this comparison *E* was aware of which Ss were role-playing and also consciously anticipated no differences between age-regressed and role-playing Ss. More relevant information on this question can be obtained from the comparison of the age-regressed and cryptosimulating groups, since *E* was blind to their identity in this situation. This is discussed below.

Hollow Tube Test. In the original study, differences were predicted at all three ages for this test. Significant differences, however, were found only at ages 10 and 7. The Ss at age 4 tended merely to say "I don't know" or the equivalent. In the present study, a significant difference was found only at age 7, although the difference at age 4 just misses being significant, the obtained *U* value being 27.5 where 27.0 is significant at the .05 level. This is an example of the fact that the best discrimination on test performance is found in both studies at age 7.

Left and Right Test. No differences in performance on this test were predicted in the original study at age 10, but a difference was expected at age 4. The present study agrees with the original one in finding no significant differences at either age, although again the *U* value at age 4 just misses being significant. The usual response at age 4 is simply "I don't know."

Word Association Test. This test shows a major failure of replication. It was one of the most discriminatory tests in the original study, showing significant differences in percentage of abstract responding at all three subadult ages. Briefly, the reason it did not show discrimination on this measure in the present study is probably due to the fact that initial levels of abstract responding were so low in the adult

waking administration that there was no room, statistically speaking, for significant drops to occur. The reaction time scores on the test do discriminate in the present study at all three subadult age levels, but comparable time data are not presented in the original study.

Arithmetic Test. This is the most consistently discriminatory test in both studies. No difference in performance level was predicted in the original study at age 10, although differences were predicted in the time measure at both ages 10 and 7.¹²

Both studies are in agreement in finding significant differences in performance level at age 7 but not at age 10, and in the time measure at both ages.

Pledge of Allegiance. A number of the measures used in the present study were either not reported in the original or not given in sufficient detail to allow exact comparisons to be made. One measure that can be compared was S's expressed ability to write. Reiff and Scheerer predicted that both experimental and control groups would be able to write at age 10 but that they would differ at the lower age. Both studies are in agreement in finding no significant differences at age 10, but they differ markedly in their results at age 7. None of Reiff and Scheerer's age-regressed Ss were able to write at this age, whereas 7 out of 10 of our age-regressed Ss could do so. They did not, in fact, differ significantly from their role-playing counterparts on this criterion.

This is a type of behavior which could be very sensitive to the amount of role support given by *E*. The expectation of *E* can be made readily apparent to S.

The occurrence of "Mondegreenisms" among groups in the present study other than the children was so slight that statistical comparisons between groups could not be performed.¹³ This is in agreement with the find-

¹² There is a minor discrepancy here. Actually, no difference is predicted for either the performance level or time measure in the text of either the thesis or the monograph. In Table J (Reiff & Scheerer, 1959, p. 174) of the monograph a significant difference in time score is indicated without explanation.

¹³ Reiff and Scheerer coined the term Mondegreenism to describe the tendency of children to substitute inappropriate words they can understand for similar sounding phrases that they do not comprehend, such as "Lady Mondegreen" for "laid him on the green."

TABLE 5

OCCASIONS OF IMPROPER WORDING ON THE PLEDGE OF ALLEGIANCE TEST THAT COULD POSSIBLY BE SCORED AS INCIDENTS OF MONDEGREENISM

Group	Age	Correct word	Written word
Age regressed	7	indivisible	"invisible"
Three-age role players	10	indivisible	"invisable"
Three-age role players	7	indivisible	"invisibil"
Child	10	which it	"wichet"
Child	10	allegiance	"a legen"
Child	10	allegiance	"a legins"
		indivisible	"in Duvisabal"
Child	7	allegiance	"of the legens"
		justice	"just us"
Child	7	allegiance	"a Legns"
		pledge	"Plag Ieay"
Child	7	allegiance	"the leiogn"
Child	7	allegiance	"the leageinc"
Child	7	allegiance	"a Legnt"
		republic	"Re Plicae"
		which it	"Wiheres"
Child	7	and to	"Unto"

ings of Reiff and Scheerer. The possible (by even a generous stretch of imagination) responses that could be categorized as Mondegreenisms are presented in Table 5. This measure is discussed again under the question of the validity of hypnotic age regression.

Two other measures on this test differentiated between groups. The inclusion of the phrase "under God" showed a significant difference at age 10, none of the age-regressed group making the inclusion, while half of the one-age role players did so. The phrase "under God" was added to the Pledge by Act of Congress in June of 1954. The behavior of the age-regressed Ss was therefore in the appropriate direction. This measure was not used by Reiff and Scheerer, not being relevant considering the ages of their Ss.

The percentage of misspellings showed a significant difference at age 4. This is in the direction predicted by Reiff and Scheerer, although the incidence of misspellings among their Ss was too low to justify analysis.

Clock Test. Differences in performance levels were predicted in the original study at both ages 7 and 4. A significant difference was found between the experimental and control groups at age 7, but none at age 4. In the present study, this test failed to differentiate at either age. At the lower age, most Ss simply say that they cannot tell time or are even unable to read the numbers.

Mud and Lollipop Test. On the assumption that 4-yr.-old children would be less fastidious than adults, the prediction was made by Reiff and Scheerer that the age-regressed Ss would accept the lollipop even though their hands were dirty from playing in the sandbox, while control Ss would reject it. This was indeed the case in the original study. In the present study, however, there was no significant difference in frequency of acceptance, all of the age-regressed group and eight of the one-age role-players group accepting the lollipop.

An important factor, again, in our results was the expectation of *E*. The lollipop was presented with confident assurance of acceptance.

Overview. An evaluation of overall differences in performance levels averaged across tests for each age by means of the ranked

TABLE 6

RESULTS OF COMPARISONS OF OVERALL PERFORMANCE LEVELS BETWEEN THE AGE-REGRESSED AND ONE-AGE ROLE-PLAYERS GROUPS BY MEANS OF THE RANKED SUMS OF RANKS TEST

Age	U	p
10	21.0	<.05
7	5.0	<.001
4	15.0	<.005

Note.—All *p* values are one-tailed.

sums of ranks procedure is given in Table 6. The fact that significant overall differences were found at all three ages argues that a significant, though not complete, replication of the findings of Reiff and Scheerer was obtained in the present experiment.

As in the original experiment, the best separation in performance seems to be at age 7, the least at age 10.

As in the original study, the Hollow Tube Test and the Arithmetic Test showed good differentiation. A major discrepancy was the failure in the present study to find differentiation with the percentage of abstract response measure of the Word Association Test. The failure to replicate the high initial levels of abstract responding found by Reiff and Scheerer suggests a probable explanation for the failure of replication.

Lack of differences in apparent behavior in the Free Play Period, as well as on such measures as ability to write the Pledge of Allegiance and acceptance of the lollipop in the Mud and Lollipop Test can be interpreted as reflecting different *E* expectations in the two experiments.

Dependence on Hypnosis

Evidence on the extent to which the differences found by Reiff and Scheerer and partially replicated in the present study are dependent upon hypnotic suggestion can be obtained by a comparison between the age-regressed and cryptosimulator groups. Additional relevant information is provided by the results of *E*'s forced-choice estimations of the group identity of Ss run blind.

These estimates were made at the termination of each experimental session, and no information feedback was given the primary *E* until the entire experiment had been com-

TABLE 7
ACTUAL AND ESTIMATED CATEGORIES OF Ss RUN BLIND
BY *E*, THE AGE-REGRESSED AND CRYPTO-
SIMULATORS GROUPS

Actual category	Estimated category		N
	Age regressed	Crypto-simulators	
Age regressed	4	6	10
Cryptosimulators	5	5	10

Note.--Fisher's exact probability test yielded nonsignificant values.

pleted. During the experiment, *E* entertained the belief that he would be able to make these estimations at better than chance accuracy. Results of his estimations, presented in Table 7, do not support this belief. A Fisher's exact probability test on this array shows no significant departure from chance. This is in keeping with previous studies using the real-simulator design (Damaser, Shor, & Orne, 1963; Evans, 1968; Orne, 1959; Orne & Evans, 1965; Orne & Evans, 1966; Sheehan, 1969; Sheehan & Orne, 1968; Shor, 1962).

Results of performance tests are in agreement with this conclusion. Only one measure, the percentage of misspellings on the Pledge of Allegiance at age 7, showed a significant difference between the age-regressed and cryptosimulator groups.

Analysis of overall performance by means of the ranked sums of ranks procedure also gave a nonsignificant *U* value of 48.0. Overall performance was evaluated by the use of tests from all three ages in this and other comparisons where the same group of Ss was given the tests for all three ages, since this increased the amount of information utilized in the evaluation.

Experimental Design Effects

Two sets of comparisons are relevant to the possible effects of the unbalanced design used in the original study, where regressed Ss were given tests at three ages while control Ss were tested at only one age.

The first comparison is between the three-age role players and the one-age role players. Since both groups were identically treated through age 10 testing, there was no a priori reason to expect any differentiation at this

age. Only one measure did show a significant difference, the reaction time to noncritical words on the Word Association Test.

At age 7, the age at which the most marked differentiation was found in both the original and present studies, the one-age role players showed an older level of performance on both the Hollow Tube Test and the Arithmetic Test. They also showed fewer misspellings on the Pledge of Allegiance, which would be predicted from a tendency to older performance.

No significant differences were found in performance measures at age 4, although there were significant differences in behavior during the Free Play Period. The one-age role players showed more reluctance to enter, no verbal interaction with *E*, and silence. These measures suggest that Ss of this group were more anxious than the three-age role players, who had had more opportunity to get into the role by this time. The Ss in both groups tended to give minimal performance on the tests. They seem to solve the problem of appropriate behavior at this age by acting as childishly as possible.

The results of the ranked sums of ranks analyses bear out the impression that the greatest differentiation is obtained at age 7. These results are presented in Table 8.

The second relevant comparison is between the age-regressed and three-age role-players groups. If hypnotic amnesia, even though present subjectively, does not prevent functional interaction from age to age, one would expect the three-age role players to match their age-regressed counterparts, at least more closely than do the one-age role players.

No differences were found between the age-regressed and three-age role players at the adult level, nor on performance measures at age 10. There were significant differences in

TABLE 8
RESULTS OF COMPARISONS OF OVERALL PERFORMANCE
LEVELS BETWEEN THE THREE-AGE ROLE-PLAYERS
AND ONE-AGE ROLE-PLAYERS GROUPS BY
MEANS OF THE RANKED SUMS
OF RANKS TEST

Age	<i>U</i>	<i>p</i>
10	27.5	<i>ns</i>
7	15.5	.005
4	32.5	<i>ns</i>

behavior during the Free Play Period, with more verbal interaction with *E* in the age-regressed group and more *Ss* in the three-age role players group playing in the sandbox. The two behaviors may be interrelated, since playing in the sandbox can be used as a maneuver to avoid talking to *E*.

Significant differences appear in two performance tests at age 7, the reaction time to critical words on the Word Association Test being longer for the age-regressed group, and both the performance level and time measures on the Arithmetic Test showing a differentiation.

At age 4 the reaction time to critical words was again significantly different and again more of the role-playing *Ss* played in the sandbox.

All in all, the number of measures that differentiate between these groups is not large. This impression is supported by the result of the ranked sums of ranks analysis, which did not indicate a significant difference in overall performance ($U = 29.0$, *ns*).

Taken together, the results of these two sets of comparisons support the hypothesis that the design used in the original study may itself have produced some of the differences found between the age-regressed and control *Ss* of Reiff and Scheerer.

Effects of Direct versus Indirect Participation

The quasi-participants were instructed to treat the testing procedures as an intellectual exercise and were told specifically that they would not participate in the experiment. In addition, *E* attempted to give minimal role support. The purpose of these instructions was to minimize the entwined factors (role support and role playing) on the part of both *E* and *S*. Comparison of this group with the three-age role players, their nearest analog in the present experiment, can provide information on the importance of role factors.

Again, no adult differences were found. At age 10 several differences appeared in the Free Play Period, more of the quasi-participating *Ss* claiming that they would talk to *E*, that they would not talk to themselves, and that they would not play in the sandbox. They were not, of course, given the opportunity to do any of these things, but were merely asked

what they thought they would do if taking part in the experiment as a regular *S*. It is difficult to interpret these apparent differences, since their statements of what they would do in the real situation may not be accurate predictors of their actual behavior. Of the performance tests, only the reaction time measures of the Word Association Test showed significant differences.

At age 7, the quasi-participants showed a significantly higher performance level on the Hollow Tube Test. Reaction time to non-critical words was again significantly faster, and the percentage of abstract responding showed a significant difference. In view of the lack of evidence for overall significance on this measure, this instance is best regarded as a sampling error. Differences persisted in their estimation of their probable behavior in the Free Play Period, although more of them felt they would not play in the sandbox.

Differences in the Free Play Period behavior became less marked at age 4, although there still seemed to be some reluctance to play in the sandbox. The reaction time to critical words was again significantly faster, but the other measures did not differentiate.

The differences found between these two groups were not as marked as expected, although the ranked sums of ranks test showed an overall difference at the $p = .05$ level ($U = 23.0$). All in all, the quasi-participants were able to approximate the behavior of the role players to a considerable degree. Either *E* was giving more role support than he realized, or *Ss* brought more to the experimental situation than *E* anticipated.

Uniformity of Experimenter Expectations

A further indication of the nature and influence of *E* expectation in this investigation is provided by a comparison of the three-age role players and cryptosimulator groups. It has been pointed out (Orne, 1959) that the instructions and demand characteristics placed upon cryptosimulating *Ss* tend to maximize their motivation, particularly in comparison with *Ss* who are asked to pretend to be hypnotized in the usual control design. In the latter, such *Ss* are very unsure of the real expectations of *E*, commonly perceiving them to indicate that the simulating *S* will be

unable to give a convincing performance, which perception of *E*'s expectations may often be quite correct. The *E* who anticipates a differential effect between his hypnotized and control groups has good opportunity to transmit this expectation covertly to *S*. Also, *S* in this situation is never sure whether he is doing an adequate role-playing job or not. It has been reported that such faking *Ss* frequently break their roles during simulation.

In the cryptosimulating situation, *E*, being blind, cannot treat his hypnotized *Ss* differently from his cryptosimulating *Ss* unless, which so far has never been the case in this type of experiment, he correctly perceives at better than chance level the true roles of *Ss* in the two groups. The design thus eliminates one potent source of differential performance effects.

Since the cryptosimulating *S* is informed that the experiment will be stopped if *E* becomes sure that he is not truly hypnotized, *S*'s motivation to continue is supported throughout the experiment as long as the experiment is not terminated by *E*.

In the present experiment, the three-age role-players groups were treated exactly as were the cryptosimulators group, with the important exception that *E* was aware, obviously, that the role players were not hypnotized. He tried to maintain an equal expectation, as has been stated above, of success for all *Ss* and to provide a maximum of role support for all *Ss*. The extent to which the cryptosimulators and three-age role-players groups produced equivalent performance may, therefore, be taken as an objective measure of the extent to which *E* was successful in maintaining constant expectations of outcome.

Individual test comparisons showed significant differences on the time measure of the Arithmetic Test at ages 10 and 7, and on the performance level of the Hollow Tube Test at age 10. The overall test of performance, however, failed to reach significance ($U = 39$, *ns*). These results support the contention that *E* was able to maintain a relatively equal attitude toward all *Ss*.

This conclusion is supported by the fact that none of the *Ss* in the three-age role-players group broke role during the experiment, nor did any in the one-age role-players

group either. This is in marked contrast to the results of Reiff and Scheerer (1959), who reported that only 2 of their 15 "simulators" even attempted to imitate the voice of a child (p. 117), whereas all their age-regressed *Ss* did so. They further stated,

Many of the control subjects found it difficult to maintain their role during the period in the playroom. Several times, either during the free-play period, or during the actual presentation of the tasks, they stepped out of their role to make some side remark to the examiner [p. 118].

No role-playing *Ss* in the present experiment ever did this.

Questionnaire Material

The original study of Reiff and Scheerer was designed primarily to test a context theory of recall by utilizing hypnotic age regression as a means of reinstating earlier modes of thinking. Hypnosis was used as an investigative tool. The present study, in contrast, is primarily concerned with the legitimacy of using this tool in this way. The assumption that hypnotic age regression can in fact reinstate earlier modes of thinking to a degree not available through appropriate instruction in the waking state was called into question. The expectation that followed from this position was that significant differences in recall would not be found among the treatment groups of the present study if necessary safeguards were taken, such as completing the entire testing in a single experimental session. The apparent hypermnnesia found by Reiff and Scheerer was regarded as an artifact. In order to test the accuracy of this position, results of analyses of the questionnaire material are now presented.

Proportion of Memories Obtained

The first measure to be considered is the proportion of credible memories obtained on successive questionnaire presentations. Credible memories are here defined as answers to the seven questions designed to evoke specific memories about the fifth grade and the additional seven such questions pertaining to the second grade. The answers to these questions were considered credible if they contained specific content. They were not considered credible if they were of the "I don't know"

TABLE 9

PROPORTION OF CREDIBLE MEMORIES
OBTAINED FROM QUESTIONNAIRES

Group	Questionnaire			
	A(wake)	A(trance)	B	C
Age regressed	.528	.662	.807	.786
Cryptosimulators	.530	.759	.804	.849
Three-age role players	.546	—	.588	.711
One-age role players				
Age 10	.514	—	.700	—
Age 7	.634	—	—	.726
Age 4	.400	—	—	—
Quasi-participants	.566	.636	.712	.726
Children				
Age 10	—	—	.822	—
Age 7	—	—	—	.986
Age 4	—	—	—	—

variety or too vague to show specific content. The mean proportions obtained from each group in the present experiment are summarized in Table 9. Differences among treatment groups (children excluded) do not appear to be marked, and this is borne out by the results of the Kruskal-Wallis analysis of variance by ranks (Siegel, 1956) which showed no significant differences among groups for each questionnaire. These results are presented in Table 10.

These results are in marked contrast to those of the original study, where in fact so few control Ss gave anything but "I don't know" or equivalent answers that no effort was made to validate from school records the few specific answers obtained.

Statistical comparisons between equivalent groups in the two studies are presented in Table 11. On all three questionnaires given to the role-playing Ss, there was significantly more apparent recall in the present study. There were no significant differences between the age-regressed groups in the two studies, however, except on the first waking presentation of Questionnaire A. This is a very interesting finding. The very low proportion of credible memories reported by Reiff and Scheerer's experimental Ss on the presentation prior to their entering hypnosis can be interpreted as a suppression effect in response to the demands of the situation, enabling these Ss to show greater apparent gains while in hypnosis. Such a suppression effect has been demonstrated before, although the conditions

TABLE 10

KRUSKAL-WALLIS ANALYSIS OF VARIANCE BY RANKS
OF PROPORTIONS OF CREDIBLE RESPONSES GIVEN
TO QUESTIONNAIRES BY TREATMENT GROUPS
(CHILDREN EXCLUDED)

Questionnaire	H	df	p
A(wake)	2.71	6	>.80
A(trance)	1.78	2	>.30
B	7.28	4	>.10
C	5.56	4	>.30

that produce the effect are not at present completely understood (Evans & Orne, 1965; London & Fuhrer, 1961; Rosenhan & London, 1963a; Rosenhan & London, 1963b; Zaman-sky, Scharf, & Brightbill, 1964). The possibility that such an effect was present in Reiff and Scheerer's experiment cannot, however, be dismissed in view of the present findings.

Validation of Specific Memories

A major methodological feature of the Reiff and Scheerer design was the attempt to validate specific memories obtained on the questionnaires either through reference to available grade school records or through the use of calendar information. As Yates (1961) has stressed, few studies in this area have taken this precaution to rule out confabulation as a source of apparent recall.

As pointed out above, no validation through school records of these items was attempted for the control Ss in the original study because of their low proportion of credible memories.

TABLE 11

COMPARISONS BETWEEN EQUIVALENT GROUPS IN THE
PRESENT STUDY AND THAT OF REIFF AND SCHEERER
OF PROPORTIONS OF CREDIBLE MEMORIES
OBTAINED FROM THE QUESTIONNAIRES

Questionnaire	Age regressed vs. experimental					
	\bar{X}_1	\bar{X}_2	U	p	N ₁	N ₂
A(wake)	.528	.270	9.5	.05	10	5
A(trance)	.662	.500	17.0	ns	10	5
B	.807	.670	12.0	ns	10	5
C	.786	.944	15.0	ns	10	5
	One-age role players vs. controls					
	\bar{X}_1	\bar{X}_2	U	p	N ₁	N ₂
A(wake)	.515	.329	124.5	.008	30	15
B	.700	.400	6.5	.02	10	5
C	.726	.314	2.5	.005	10	5

Validation of school data was therefore available only for the five age-regressed Ss.

The first of the items to be validated was the day of the week on which previous birthdays had fallen. This is particularly important data since it allows a uniform test, independent of the availability of school records, to be made for all Ss who give credible answers on these items.

There are good reasons for expecting valid hypermnnesia on this type of item during age regression. Particular care was taken to allow adequate induction time for the consolidation of age regression and to identify the hypnotist as someone familiar to and liked by S as a child. Again, the general emotional importance of birthdays to most children should facilitate recall.

Detailed analyses of accuracy of day-of-the-week information and of specific items from the questionnaires are available in Appendix A (see Footnote 11).

"I don't know" and equivalent answers were excluded from the calculation of overall percentages. The small cell frequencies make exact statistical analysis impractical; however, inspection shows no marked differences among groups. The overall percentages of accuracy certainly do not exceed the expected chance accuracy of 14.3%.

There are difficulties in comparing these results with those of Reiff and Scheerer. They did not, for example, give data on the accuracy of answers on the waking and trance administrations of Questionnaire A. Reiff and Scheerer (1959) summarized their results as follows:

At regressed age *ten*, out of five experimental subjects, three named the wrong day, one gave an "I

don't know" answer, and one was correct. At regressed age *seven*, one was incorrect, one said, "I don't know," and three correctly named the day of the week of their seventh birthday. One of the ten controls gave a correct answer but said later that he had just guessed. The other nine either guessed incorrectly or said, "I don't know" [p. 193].

Again, the frequencies are too small to make exact comparisons with our results. The results of both experiments, however, differ markedly from those first reported by True (1949), who reported 93% accuracy at age 10, 82% at age 7, and 69% at age 4. True's results, as Barber (1962) has pointed out, have not been replicated by other investigators, the important variable apparently being that Ss were tested intermittently over periods of months in the original study. This factor was also present to some extent in Reiff and Scheerer's design, since all three ages were not tested during a single experimental session.

An attempt was then made in the original study to use available school records to validate the following information: (a) the name of the school or schools attended during the second and fifth grades; (b) the name of the second-grade teacher; (c) the name of the fifth-grade teacher; and (d) the names of two classmates in each of these classes.

The difficulty of obtaining such information was stressed by Reiff and Scheerer and amply confirmed by our own experience. All five of their experimental Ss had attended the Topeka grade school system, which greatly aided validation. In the present sample, the geographical range of schools attended covered most of the continental United States and Hawaii as well. This did little to facilitate validation.

TABLE 12
SUMMARY OF RESULTS OF VALIDATION OF SPECIFIC MEMORIES BY REIFF AND SCHEERER

Subject	School and years	Fifth-grade teacher	Fifth-grade classmates	Second-grade teacher	Second-grade classmates
Age regressed					
A	yes	yes	yes, yes	—	—
B	yes	yes	yes, yes?	yes	yes, no
C	yes	yes	yes, no	yes	yes, yes
D	yes	yes	yes, —	yes	yes, yes
E	yes	yes	yes, no	yes	yes, yes
Total	5/5 (100%)	5/5 (100%)	7/9 (78%)	4/4 (100%)	7/8 (87.5%)
Awake					
Total	?	5/5 (100%)	2/9 (22%)	0/4? (0%)	0/8 (0%)

TABLE 13

FREQUENCIES AND PERCENTAGES (IN PARENTHESES) OF
VALIDATED CREDIBLE ANSWERS TO SPECIFIC
MEMORY ITEMS ON QUESTIONNAIRES
AVERAGED ACROSS GROUPS

Memory item	Questionnaire form		
	A (wake)	A (trance)	B or C
Schools attended			
Fifth grade	29/50 (58%)	16/30 (53%)	B. 30/50 (60%)
Second grade	23/49 (47%)	12/30 (40%)	C. 26/50 (52%)
Teachers' names			
Fifth grade	11/17 (65%)	6/10 (60%)	B. 10/18 (56%)
Second grade	11/16 (69%)	2/6 (33%)	C. 12/18 (67%)
Classmates' names			
Fifth grade	10/14 (71%)	7/9 (78%)	B. 17/26 (65%)
Second grade	2/4 (50%)	2/4 (50%)	C. 12/23 (52%)

A summary of Reiff and Scheerer's results, as far as they can be determined from the tables and text of the monograph, is given in Table 12. The first *S* entered the Topeka school system in the third grade; therefore no validation of the second-grade information was made (since only Topeka school system records were consulted). The fourth *S* stated that no one sat behind him in the fifth grade. (In the present experiment, he would have been asked to supply the name of a classmate who sat beside him, since the purpose of the question was to elicit the names of two fellow classmates, regardless of their seating position.)

All in all, the amount of accurate recall during regression is most impressive, particularly when contrasted with waking performance. Detailed data for the waking administration of Questionnaire A were not given by Reiff and Scheerer, nor was any information given on its administration during hypnosis but prior to age regression. This would be of particular interest, since there is a marked increase in credible memories for these *Ss* from waking to trance presentation. The difference in accurate recall demonstrated in these data must, again, be viewed in terms of a possible suppression effect in the waking state, as well as in terms of possible effects of intervening recall processes during the time between experimental sessions.

For the present study, validation data of credible memories averaged across groups (children omitted) are presented in Table 13. The percentage of validated memories is

TABLE 14

INCIDENCE OF TRANCE-LIKE ROLE INVOLVEMENT
REPORTED BY THREE-AGE ROLE PLAYERS
AND ONE-AGE ROLE PLAYERS

Trance-like involvement	Role-playing group				Total
	Three-age	One-age 10	One-age 7	One-age 4	
Experienced	5	2	4	6	17 (42.5%)
Not experienced	5	8	6	4	23 (57.5%)

not as high as that obtained by Reiff and Scheerer's experimental *Ss*, but nevertheless is quite substantial.

The major differences between the two experiments are the lack of any marked difference in proportion of credible memories or proportion of validated memories between the age-regressed group and other groups, and the lack of an apparent suppression effect on the waking presentation of the questionnaire material.

Other Findings

Reports of trance-like role involvement. The three-age role players and the one-age role players were questioned during the postexperimental inquiry about the spontaneous occurrence, while they were acting out the role of a child, of feelings of being in some way back at the age they were instructed to role-play. Such subjective occurrences could be interpreted as trance-like in quality, analogous to hypnotically suggested age regression. They were asked, specifically, "Did you at any time feel that you were (age) yr. old?" The incidence of reports of such occurrences for these groups is presented in Table 14.

The proportion of *Ss* who reported feelings of this type is rather surprisingly high. Most reports of these trance-like experiences indicated that they occurred during the Free Play Period or during the Mud and Lollipop Test.

The occurrence of these reports was not found to be a function of the hypnotizability of *Ss*, as later tested with the Stanford scale. The mean for *Ss* giving positive reports was 7.1 and the standard deviation was 3.6; those for *Ss* giving negative reports were 6.2 and 2.4. These distributions do not differ significantly ($t = .87, ns$).¹⁴

¹⁴ Corrected for heterogeneity of variance.

The age-regressed Ss, as would be expected, reported subjective conviction of being at the suggested age throughout the experiment. None of the role-playing Ss described experiences that were comparable in quality and intensity to the experienced reality of the age-regressed group.

Strong role support was given to these role-playing Ss. It is of interest to look, therefore, at similar reports from the quasi-participants, who were given minimal role support. Two of the Ss in this group gave positive reports to the question of realistic role involvement. It should be pointed out again that these Ss were instructed to regard the experimental situation as a dry intellectual exercise. It is somewhat surprising that even under these pallid conditions, reports of spontaneous trance-like experiences were obtained.

Behavioral validity of age regression. A basic question in the history of hypnotic research has been the validity of age-regression phenomena. This question may be asked in a variety of ways. For example, one may question the subjective validity, the (neuro-) physiological validity, or the behavioral validity of the phenomenon. The general consensus agrees on the subjective reality of age-regression phenomena, although this often shows fluctuation and is complete only in very good hypnotic Ss. Evidence for age regression in any true neurophysiological or physiological sense is scant and most controversial. The evidence for behavioral regression is considerably better, although most investigators have found that the behavior produced is somewhat older than that appropriate to the age suggested. These generalizations are based on the reviews of the field by Barber (1962), Gebhard (1961), and Yates (1961). A recent critical discussion of this material can be found in Hilgard (1965).

In order to get some indication of the validity of the age-regressed Ss in the present study, three groups of children were run under as nearly as possible identical conditions at ages 10, 7, and 4. Although these groups of children are here designated validation groups, we are under no illusions about the generality of the behavioral measures obtained from them. The samples are obviously too small and too unrepresentative to provide

more than a tachistoscopic look at what real children do under our testing conditions. There follows a description of comparisons of the age-regressed group with actual children.

Four measures showed significant differences at age 10. The reaction time to critical words was significantly shorter for the children than for the age-regressed Ss. This could be due to lack of affective loading on these words for the children, although as a matter of fact many of them were quite aware of the meanings of these words. There was a considerable difference in the time measures of the Arithmetic Test. The age-regressed Ss took almost twice as long to solve these problems as did the 10-yr-olds. In a sense, the age-regressed group was acting more childlike than the real children. The children, on the other hand, showed more misspellings on the Pledge of Allegiance. Here the regressed adults were performing at too high a level. Seven of the eight children who completed the Pledge included the phrase "under God," whereas none of the age regressed did so. Both groups, however, were behaving as they should, since the phrase was not part of the Pledge during the years when the hypnotic Ss actually were 10 yr. old.

The only test showing significant differences at age 7 was the Arithmetic Test. The children were performing at a higher level and taking less time to do so than were the regressed Ss.

At age 4, only the performance level on the Hollow Tube Test differentiated between the groups. The regressed Ss were performing at a higher level than the children. It is interesting to note again that children of this age did accept the lollipop on the Mud and Lollipop Test.

All in all, there seems to be little evidence as far as these behavioral measures are concerned that the age-regressed group was not behaving appropriately at these three ages. Results of analysis by the ranked sums of ranks procedure are presented in Table 15, where only at age 7 was there an overall difference in performance, and probably this small difference is attributable almost entirely to the Arithmetic difference.

The performance tests, however, do not tell the whole story. There are marked qualitative

differences in behavior, difficult to quantify. An example has already been given in the incidence of Mondegreenisms in the children group and their comparative absence in the adult groups.

Another and very striking difference was found in administering the Word Association Test to the age-4 children. The tendency to concrete response was here carried to an extreme. Many of the children gave not only a verbal response but also acted it out in some way, or gave only a nonverbal response. For example, in response to the stimulus word RED, the child might hand *E* a red toy, saying "Like this!" or some equivalent. None of the adult Ss showed this type of acting-out response.

DISCUSSION

The implications of the present experimental findings primarily fall into two categories. First, substantive questions concerning the validity of hypnotic age regression, and the apparent hypermnnesia obtained therefrom, are raised. The relation of the present findings to the interpretation of those of Reiff and Scheerer, in particular, falls into this category. Second, a number of methodological questions for research in hypnosis in general are raised. These questions are obviously interrelated.

Reiff and Scheerer observed consistent dramatic and significant differences between the age-regressed Ss and their one-age role-playing control Ss. Based on these observations, they felt that the reinstatement of prior functional schemata was possible through hypnotic age regression.

In our replication, the behavior of the age-regressed Ss, on the whole, effectively replicated that reported by Reiff and Scheerer. We are satisfied that the phenomenon of hypnotic age regression observed in our laboratory was the same as that observed by Reiff and Scheerer. In the comparison between the behavior of these age-regressed Ss and one-age role players, our results were the same in some aspects, but there were also important failures of replication which do not seem to be merely instances of sampling fluctuation. These differences, however, were in the behavior of the

TABLE 15
RESULTS OF COMPARISONS OF OVERALL PERFORMANCE
LEVELS BETWEEN THE AGE-REGRESSED AND
CHILDREN GROUPS BY MEANS OF THE
RANKED SUMS OF RANKS TEST

Age	<i>U</i>	<i>p</i>
10	36.0	<i>ns</i>
7	23.0	<.05
4	30.0	<i>ns</i>

role players, *not* in the behavior of the age-regressed Ss. For example, our role-playing Ss did not break out of character, as did those of Reiff and Scheerer, nor did they show overt signs of nervousness, such as defensive laughing, as did their counterparts in the original study. Equally striking is the apparent ability of the one-age role players to write the Pledge of Allegiance Test and their willingness in the present experiment to accept the lollipop in the Mud and Lollipop Test.

While in the original study, consistent significant differences were observed between age-regressed Ss and one-age role players, our one-age control Ss, while still different, were much more capable of behaving as the regressed Ss did. In the original study, for example, all the one-age role players at age 7 reported that they could not tell time, thus yielding a significant difference between age-regressed and control groups. This was not the case in the present study. Particularly striking is the difference in the behavior of our control Ss in the proportion of credible memories obtained with the questionnaire. Reiff and Scheerer obtained so few appropriate memories from the control group that they did not justify an attempt at validation, whereas we observed no significant difference of this kind.

It is interesting to inquire into the reasons for the differences between the behavior of Reiff and Scheerer's one-age role players and ours. It is our impression that *E*, by virtue of his expectations and his conscious effort, gave more role support to the one-age role players, and it was in those procedures which were open to the influence of such factors that the failure of replication occurred. It may be added that *E* bias is not necessarily conscious. Indeed, Rosenthal's (1963) results imply that

its effect is weakened when *E* is induced consciously to bias his work.

Some of our results (e.g., the Arithmetic and Hollow Tube Tests) agreed with those of Reiff and Scheerer; however, even if *all* the comparisons of these groups had been in complete agreement, we would still question their conclusions. We recognized that Reiff and Scheerer's controls were inadequate to the task in a number of ways (Orne & O'Connell, 1961). Implicit in their conclusions were the assumptions (*a*) that the behavioral tasks employed were such that they could not be mimicked successfully by control Ss; (*b*) that there were no significant differences in the treatments of age-regressed and control Ss in regard to role support or subtle bias; and (*c*) that the control Ss would be truly motivated to give their best performance in attempting to behave as children.

Our additional control and comparison groups were designed in a variety of ways to test these assumptions and determine the extent to which alternative mechanisms could more parsimoniously account for Reiff and Scheerer's data. The body of evidence indicates that the predictions of our study have been substantially confirmed, leading us to conclude that Reiff and Scheerer were mistaken and that hypnotically age-regressed behavior does not involve the total reinstitution of childlike mental processes and associated memories. Where activities were shown through controls to be in some way amenable to influence, such as *E* support, demand characteristics, and heightened motivation, age-regressed Ss were able to perform in a childlike way. Where the activities were less amenable to such influence, age-regressed Ss performed no more successfully than the other groups.

The main body of evidence for the above conclusion consists of three interrelated parts. First, three-age role players are in a better position to produce an appropriately graduated series of performance than three groups of one-age role players, and our results show that they take advantage of it. Second, when the most important comparison group (the cryptosimulators) was added, insuring that *E* provide identical cues, demands, and supports

to the comparison group and to the age-regressed group, and that *S* do his best to produce the desired behavior, then differences between regressed and nonregressed Ss are no longer found. Clearly the behavioral indicants of hypnosis used in this study, contrary to the assumptions of Reiff and Scheerer, are simulable by un hypnotized Ss under suitable circumstances. The "genuineness" or the "fullness" of regression thus remains unproved. Third, most striking are the various demonstrations of features in the behavior of real children which neither regressed Ss nor cryptosimulators effectively reproduced, for example, the acting-out response on the Word Association Test. It is surprising that with the exception of the studies of Sarbin (1950) and Orne (1951), no attempts have been made to compare actual children's responses with those of age-regressed Ss. When such comparisons are made, the qualitative differences between the age-regressed *S* and the child dwarf those observed between age-regressed Ss and *any* control group.

We had initially anticipated that our one-age role players would behave similarly to those of Reiff and Scheerer, not fully recognizing the effect the different expectations of the investigator might have on these control Ss. Had the behavior of the one-age role players been the same in both studies, the differences between the one-age and the three-age role players would have been more pronounced than was the case and we would most likely have observed clear differences between the three-age role players and the cryptosimulators. However, in the present study few such differences were noted, but it should not, therefore, be concluded that the absence of striking differences between the three-age role players and the cryptosimulators would make the latter group unnecessary in another study. Indeed, we believe that if *E* had expected large differences between the age-regressed and the control Ss, we would have found major differences between role players and cryptosimulators. The use of cryptosimulators (or a variant of this technique as in the London and Fuhrer (1961) design) is the only way one can be certain that the investigator will treat experimental and control Ss in the identical

manner. Obviously, if no differences are found between role-playing controls and hypnotized Ss, the cryptosimulator groups are unnecessary. However, once they are found, the cryptosimulators become an essential comparison group.

It does not follow from the above that age-regressed Ss are consciously simulating. Post-experimental inquiries revealed profound subjective alterations as the result of hypnotic suggestions, although *E* could not (without access to the inquiry material) distinguish at better than chance levels between regressed and cryptosimulating Ss on the basis of overt behavior. The mechanisms by which these behavioral patterns are produced are not the same.

Other evidence attests to the genuineness of hypnotic phenomena (see Orne, 1970). For example, in an experiment on hypnotic amnesia which did make use of a group of simulators, Williamsen, Johnson, and Eriksen (1965) were led to conclude that

posthypnotic amnesia impaired recall and recognition among hypnotized Ss, but did not reduce the availability of the words as associative responses. The simulating Ss over-played their amnesic role and also showed impaired performance on the associative tests [p. 123].

The findings of Williamsen, Johnson, and Eriksen, and those of Leonard (1965) referred to earlier, point to a repression model of posthypnotic amnesia rather than to one of either functional ablation or of simulation, and our results are consistent with that position.

Similarly, Orne (1951) described an individual age regressed to age 6 printing in a childlike fashion but with perfect spelling, "I am conducting an experiment which will assess my psychological capacities [p. 219]." Instances such as these speak against the concept of functional ablation while, at the same time, they strongly support the subjective reality of the experience. It would seem unlikely that a role player would have permitted such incongruity in his behavior.

We do not believe that other hypnotic phenomena are to be viewed as nongenuine because of this finding about age regression. For example, another phenomenon closely linked

with regression is revivification, in which *S* appears again to live through previous experiences. As is well-known, this kind of regression is believed to be of great clinical significance. While in age regression *S* is instructed to return to a specific time, in revivification *S* is free, more or less spontaneously, to return to some meaningful past experience. Age regression, as studied in the laboratory, is rarely accompanied by profound emotional experiences or overt evidence of extreme affect, but revivification in a therapeutic context recaptures experiences that evoke extreme feeling states, almost invariably frightening and extremely unpleasant for the patient. Typically, some measure of spontaneous amnesia for these events occurs when the patient awakens from his experience. Neither the therapeutic importance nor the genuineness of revivification is challenged by our present findings. Certainly the requirements for its occurrence may well be different. While it is likely that it shares some characteristics of age regression, it is by no means identical to it.

A serendipitous finding deserving brief comment was the smallness of the differences between the quasi-participants and the role players, particularly in the light of conclusions by Troffer (1965) that

The clearest and most consistent result of the study is the superiority of high role support over low role support in producing child-appropriate responses on the cognitive tests [p. 87].

Possibly in our study, the quasi-participants were more clearly cast in the role of co-*E* (Orne, 1969), and so were made more independent of role than was anticipated.

Two methodological matters are raised by these studies: one, the matter of the types of control or quasi-control groups in investigations of hypnotic phenomena; the other, the matter of the suitability of using hypnosis as a tool in the investigation of nonhypnotic mental processes.

Two factors of particular importance in experimental hypnotic investigations are *E* bias and the demand characteristics of the situation (Orne, 1959, 1962, 1969). Several procedures designed to control or ascertain

the effects of these factors were included in the present experiment. First, the primary *E* strove to maintain a conscious expectation of equal performance by all adult groups with the obvious exception of the quasi-participants, thus reducing *E* bias. Second, three quasi-control procedures were used to deal with demand characteristics (i.e., *S*'s perception of the purpose and desired outcome, his intuitions about what is "right" in the situation). One was the post-experimental inquiry, in which *S* became, in some sense, a co-*E*, helping to discover what was actually going on as he reacted in the situation. A second was the use of quasi-participants,¹⁵ similar in some ways to the inquiry, but giving responses that are not so clearly retrospective and which are in the same form as the responses of the real experimental group. The third, and perhaps most powerful, was the use of cryptosimulators. Their behavior provided an excellent estimate of the extent to which demand characteristics and role taking were able to produce behavior comparable to that produced ostensibly by hypnotic suggestion. It showed that responses thought to be difficult or impossible of simulation were not so.

Whether hypnotically induced phenomena can be used effectively in the study of non-hypnotic mental processes is a moot point. For example, it is uncertain whether ordinary dreams and hypnotic dreams are sufficiently alike for the latter to be viewed as substitutes for the former in experimental work. Attempts have been made with some success to use hypnosis to bring about repression (Bobbitt, 1958) and anxiety (Levitt, Persky, & Brady, 1964). What our study has demonstrated is that the equivalence of waking and hypnotic phenomena which seem the same cannot simply be taken for granted. The matter must be empirically investigated in each instance. The investigator who is considering the use of hypnosis as a manipulative procedure for the investigation of broader general psychological processes should initiate his study with hesitancy, execute it with caution, and interpret its results with reservation.

¹⁵ This type of group has been described elsewhere as constituting a "non-experiment" or "pre-inquiry" (Orne, 1969).

REFERENCES

- ASHLEY, W. R., HARPER, R. S., & RUNYON, D. L. The perceived size of coins in normal and hypnotically induced economic states. *American Journal of Psychology*, 1951, 64, 564-572.
- AUBLE, D. Extended tables for the Mann-Whitney statistic. *Bulletin of the Institute for Educational Research of Indiana University*, 1953, 1, 1-39.
- BARBER, T. X. Hypnotic age regression: A critical review. *Psychosomatic Medicine*, 1962, 24, 286-299.
- BOBBITT, R. A. The repression hypothesis studied in a situation of hypnotically induced conflict. *Journal of Abnormal and Social Psychology*, 1958, 56, 204-212.
- BONEAU, C. A. A comparison of the power of the *U* and *t* tests. *Psychological Review*, 1962, 69, 246-256.
- BRADLEY, J. V. Distribution-free statistical tests. (USAF WADD Tech. Rep. No. 60-661) Washington, D. C.: United States Government Printing Office, 1960.
- BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology*, 1947, 42, 33-44.
- DAMASER, E. C., SHOR, R. E., & ORNE, M. T. Physiological effects during hypnotically requested emotions. *Psychosomatic Medicine*, 1963, 25, 334-343.
- EVANS, F. J. Recent trends in experimental hypnosis. *Behavioral Science*, 1968, 13, 477-487.
- EVANS, F. J., & ORNE, M. T. Motivation, performance, and hypnosis. *International Journal of Clinical and Experimental Hypnosis*, 1965, 13, 103-116.
- GAKKEBUSH, V. M. The use of hypnotic inhibition to study the development of human personality. *The Johns Hopkins University Applied Physics Laboratory, Library Bulletin (Translations Series)*, Report No. TG 230-T153, 1960. (Translation from *Sovremennaya Psikhonevrologiya*, 1928, 7, 272-277.)
- GAKKEBUSH, V. M., POLINKOVSKII, S. I., & FUNDILER, R. I. Experimental study of personality development by hypnotic inhibition. *The Johns Hopkins University Applied Physics Laboratory, Library Bulletin (Translations Series)*, Report No. TG 230-T152, 1960. (Translation from *Trudy Instituta Psikhonevrologiia (Kiev)*, 1930, 2, 236-272.)
- GEBHARD, J. W. Hypnotic age-regression: A review. *American Journal of Clinical Hypnosis*, 1961, 3, 139-168.
- HILGARD, E. R. *Hypnotic susceptibility*. New York: Harcourt, Brace & World, 1965.
- HULL, C. L. *Hypnosis and suggestibility: An experimental approach*. New York: Appleton-Century-Crofts, 1933.
- KENDALL, M. G. *Rank correlation methods*. London: Ch. Griffin, 1948.
- KENDALL, M. G. *Rank correlation methods*. (2nd ed.) London: Ch. Griffin, 1955.
- KLINE, M. V. An hypnotic experimental approach to the genesis of occupational interests and choice: II. The Thematic Apperception Test (a case report). *Journal of General Psychology*, 1953, 48, 79-82.

- KLINE, M. V., & HAGGERTY, A. D. An hypnotic experimental approach to the genesis of occupational interests and choice: III. Hypnotic age regression and the Thematic Apperception Test—a clinical case study in occupational identification. *Journal of Clinical and Experimental Hypnosis*, 1953, 1, 18-31.
- KLINE, M. V., & SCHNECK, J. M. An hypnotic experimental approach to the genesis of occupational interests and choice: I. Theoretical orientation and hypnotic scene visualization. *British Journal of Medical Hypnotism*, 1950, 2, 1-10.
- LECROON, L. M. A study of age regression under hypnosis. In L. M. LeCron (Ed.), *Experimental hypnosis: A symposium of articles on research by many of the world's leading authorities*. New York: MacMillan, 1948.
- LEONARD, J. R. Hypnotic age regression: A test of the functional ablation hypothesis. *Journal of Abnormal Psychology*, 1965, 70, 266-269.
- LEVITT, E. E., PERSKY, H., & BRADY, J. P. *Hypnotic induction of anxiety: A psychoendocrine investigation*. Springfield, Ill.: Charles C Thomas, 1964.
- LONDON, P., & FUHRER, M. Hypnosis, motivation and performance. *Journal of Personality*, 1961, 29, 321-333.
- MEARES, A. *A system of medical hypnosis*. Philadelphia: Saunders, 1960.
- O'CONNELL, D. N. Hypnotic age regression. Paper presented at the meeting of the Society for Clinical and Experimental Hypnosis, Cleveland, October 1961.
- O'CONNELL, D. N., ORNE, M. T., & SHOR, R. E. A comparison of hypnotic susceptibility as assessed by diagnostic ratings and initial standardized test scores. *International Journal of Clinical and Experimental Hypnosis*, 1966, 14, 324-332.
- ORNE, M. T. The mechanisms of hypnotic age regression: An experimental study. *Journal of Abnormal and Social Psychology*, 1951, 46, 213-225.
- ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology*, 1959, 58, 277-299.
- ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, 17, 776-783.
- ORNE, M. T. Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. Rosnow (Eds.), *Artifact in behavioral research*. New York: Academic Press, 1969.
- ORNE, M. T. Hypnosis, motivation and the ecological validity of the psychological experiment. Paper read at the Nebraska Symposium on Motivation, Lincoln, March 1970. In *Nebraska Symposium on Motivation*, 1970, in press.
- ORNE, M. T., & EVANS, F. J. Social control in the psychological experiment: Antisocial behavior and hypnosis. *Journal of Personality and Social Psychology*, 1965, 1, 189-200.
- ORNE, M. T., & EVANS, F. J. Inadvertent termination of hypnosis with hypnotized and simulating subjects. *International Journal of Clinical and Experimental Hypnosis*, 1966, 14, 61-78.
- ORNE, M. T., & O'CONNELL, D. N. Age regression by hypnosis. Review of R. Reiff & M. Scheerer, *Memory and hypnotic age regression: Developmental aspects of cognitive function explored through hypnosis*. *Contemporary Psychology*, 1961, 6, 70-72.
- ORNE, M. T., & O'CONNELL, D. N. Diagnostic ratings of hypnotizability. *International Journal of Clinical and Experimental Hypnosis*, 1967, 15, 125-133.
- REIFF, R. Memory in hypnotic age regression. Unpublished doctoral dissertation, University of Kansas, 1954.
- REIFF, R., & SCHEERER, M. *Memory and hypnotic age regression: Developmental aspects of cognitive function explored through hypnosis*. New York: International Universities Press, 1959.
- ROSENHAN, D., & LONDON, P. Hypnosis: Expectation, susceptibility, and performance. *Journal of Abnormal and Social Psychology*, 1963, 66, 77-81. (a)
- ROSENHAN, D., & LONDON, P. Hypnosis in the un-hypnotizable: A study in rote learning. *Journal of Experimental Psychology*, 1963, 65, 30-34. (b)
- ROSENTHAL, R. On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *American Scientist*, 1963, 51, 268-283.
- ROSENTHAL, R. *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts, 1966.
- SARBIN, T. R. Mental age changes in experimental regression. *Journal of Personality*, 1950, 19, 221-228.
- SHEEHAN, P. W. The artificial induction of post-hypnotic conflict. *Journal of Abnormal Psychology*, 1969, 74, 16-25.
- SHEEHAN, P. W., & ORNE, M. T. Some comments on the nature of posthypnotic behavior. *Journal of Nervous and Mental Disease*, 1968, 146, 209-220.
- SHOR, R. E. Physiological effects of painful stimulation during hypnotic analgesia under conditions designed to minimize anxiety. *International Journal of Clinical and Experimental Hypnosis*, 1962, 10, 183-202.
- SHOR, R. E., ORNE, M. T., & O'CONNELL, D. N. Psychological correlates of plateau hypnotizability in a special volunteer sample. *Journal of Personality and Social Psychology*, 1966, 3, 80-95.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- STRICKLER, C. B. A quantitative study of post-hypnotic amnesia. *Journal of Abnormal and Social Psychology*, 1929, 24, 108-119.
- TROFFER, S. A. Hypnotic age regression and cognitive functioning. Unpublished doctoral dissertation, Stanford University, 1965.
- TRUE, R. M. Experimental control in hypnotic age regression states. *Science*, 1949, 110, 583-584.
- TRUE, R. M. Limitations of hypnotic behavior. Panel discussion presented at the New England Society

- of Clinical Hypnosis Workshop, Weston, Massachusetts, May 1962.
- WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Forms A and B*. Palo Alto, Calif.: Consulting Psychologists Press, 1959.
- WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Form C*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.
- WILLIAMSEN, J. A., JOHNSON, H. J., & ERIKSEN, C. W. Some characteristics of posthypnotic amnesia. *Journal of Abnormal Psychology*, 1965, **70**, 123-131.
- YATES, A. J. Hypnotic age regression. *Psychological Bulletin*, 1961, **58**, 429-440.
- ZAMANSKY, H. S., SCHARF, B., & BRIGHTBILL, R. The effect of expectancy for hypnosis on prehypnotic performance. *Journal of Personality*, 1964, **32**, 236-248.

(Received August 15, 1968)